

Identification of novel tumor predisposition families and underlying genetic defects

Mervi Aavikko, MSc

Department of Medical Genetics
Haartman Institute
&
Genome-Scale Biology
Research Programs Unit
Faculty of Medicine
University of Helsinki
Helsinki, Finland

Helsinki Graduate Program in Biotechnology and Molecular Biology (GPBM)
and Integrative Life Science (ILS) Doctoral Program

Academic dissertation

*To be publicly discussed, with the permission of the Faculty of Medicine, University of Helsinki,
Haartman Institute, Lecture hall 1, Haartmaninkatu 3, Helsinki,
17th of October 2014, at 12 noon*

Helsinki 2014

Supervised by Academy Professor Lauri A. Aaltonen, MD, PhD
Department of Medical Genetics
Research Programs Unit, Genome-Scale Biology
Haartman Institute
University of Helsinki
Helsinki, Finland

Docent Pia Vahteristo, PhD
Department of Medical Genetics
Research Programs Unit, Genome-Scale Biology
Haartman Institute
University of Helsinki
Helsinki, Finland

Reviewed by Professor Samuli Ripatti, PhD
Hjelt Institute and Institute for Molecular Medicine Finland (FIMM)
University of Helsinki
Helsinki, Finland
National Institute for Health and Welfare, Unit of Public Health Genomics
Helsinki, Finland
Wellcome Trust Sanger Institute
Hinxton, Cambridge, United Kingdom

Docent Minna Männikkö, PhD
Institute of Health Sciences
University of Oulu
Oulu, Finland

Official opponent Professor Kari Hemminki, MD, PhD
German Cancer Research Center (DKFZ)
Division of Molecular Genetic Epidemiology
Heidelberg, Germany
Center for Primary Health Care Research
Lund University
Malmö, Sweden

ISBN 978-951-51-0252-2 (paperback)

ISBN 978-951-51-0253-9 (PDF)

<http://ethesis.helsinki.fi/>

Unigrafia Oy

Helsinki 2014

Table of Contents

LIST OF ORIGINAL PUBLICATIONS	5
ABBREVIATIONS	6
ABSTRACT	7
REVIEW OF THE LITERATURE.....	9
1. Tumorigenesis	9
2. Human genome.....	11
3. Tumor genome	13
3.1 Driver mutations and genes	15
4. Inherited susceptibility to tumors.....	16
4.1 Rare high risk variants	17
4.2 Rare moderate risk variants	18
4.3 Common low risk variants.....	18
5. Infectious causes of cancer	19
6. Identification of rare high risk tumor susceptibility genes	22
6.1 Linkage analysis	22
6.2 High-throughput sequencing	23
7. Tumor susceptibility conditions scrutinized in this study.....	24
7.1 Kaposi sarcoma.....	24
7.2 Nodular lymphocyte predominant Hodgkin lymphoma	25
7.3 Meningioma.....	26
AIMS OF THE STUDY	28
MATERIALS AND METHODS.....	29
1. Ethical considerations.....	29
2. Study subjects and samples.....	29
2.1 Patients in the Finnish Cancer Registry (I).....	29
2.2 Kaposi sarcoma patients (II).....	29
2.3 Hodgkin lymphoma patients (III)	30
2.4 Meningioma patients (IV)	31
2.5 Control samples (II, III, and IV).....	32
2.6 DNA and RNA extractions (II, III and IV) and cDNA synthesis (II and III)	32
3. Clustering of the patients in the Finnish Cancer Registry (I).....	33
3.1 Clustering procedure (I).....	33
3.2 Cluster score calculation (I).....	33
4. Genetic analyses	33
4.1 Genotyping (II, III and IV)	33
4.2 Transcriptome analysis (III and IV)	34
4.3 DNA copy number analysis (IV).....	34
4.4 Linkage analysis (II, III, and IV)	34
4.5 Exome and genome sequencing (II, III, and IV)	35
4.6 PCR and Sanger sequencing (II, III, and IV)	37
4.7 Loss of heterozygosity analysis (IV)	37
5. Protein analyses.....	37
5.1 Serological testing of human herpes virus 8 (I and II)	37
5.2 <i>In silico</i> predictions and modelling (II, III and IV)	37

5.3	Flow cytometry (II)	37
5.4	Cloning and site-directed mutagenesis (IV)	38
5.5	Cell culturing and production of ShhN conditioned media (IV)	39
5.6	Luciferase assay (IV)	39
5.7	Immunoprecipitation and western blot (IV)	39
5.8	Immunocytochemistry (IV)	39
RESULTS		40
1.	Cancer registry based familial clustering of cancers in Finland	40
1.1	Frequent familial clustering of Kaposi sarcoma in Finland	40
2.	STAT4 as a candidate susceptibility gene for Kaposi sarcoma	42
2.1	Prioritization of the candidate variants	43
2.2	STAT4 activation is not affected by p.Thr446Ile mutation	44
2.3	IFN- γ production is decreased in the p.Thr446Ile variant carriers	44
2.4	IL-4 and IL-17 responses in the p.Thr446Ile carriers	44
2.5	Screening of <i>STAT4</i> and associated genes	44
3.	NPAT may predispose to nodular lymphocyte predominant Hodgkin lymphoma	45
3.1	Analysis of large chromosomal alterations and EBV-infection	45
3.2	Identification of <i>NPAT</i> as a candidate susceptibility gene	45
3.3	Expression of <i>NPAT</i> in the c.2437-2438delAG carriers	46
3.4	In-frame deletion of p.Ser725 in Hodgkin lymphoma	47
4.	Loss-of-function mutation in SUFU in meningioma predisposition	47
4.1	<i>NF2</i> is not associated with the Finnish meningioma family	47
4.2	Identification of meningioma susceptibility genes	48
4.3	<i>In silico</i> modelling of the p.Arg123Cys SUFU mutation	50
4.4	Mutated <i>SUFU</i> leads to dysregulated Hedgehog signaling	51
4.5	Screening of <i>SUFU</i> in additional meningioma samples	52
DISCUSSION		54
1.	Familial aggregation of tumors in Finland	54
1.1	Identification of potential tumor susceptibility families	56
2.	Kaposi sarcoma in the Finnish population	57
2.1.	<i>STAT4</i> and Kaposi sarcoma	58
3.	NPAT in Hodgkin lymphoma predisposition	60
4.	SUFU, Hedgehog signaling, and meningioma	62
CONCLUDING REMARKS AND FUTURE PROSPECTS		65
ACKNOWLEDGEMENTS		68
WEB-BASED RESOURCES		70
REFERENCES		71

LIST OF ORIGINAL PUBLICATIONS

The thesis is based on the following publications which are referred to in the text by Roman numerals I-IV.

- I Kaasinen E^{*}, **Aavikko M^{*}**, Vahteristo P, Patama T, Li Y, Saarinen S, Kilpivaara O, Pitkänen E, Knekt P, Laaksonen M, Artama M, Lehtonen R, Aaltonen LA, Pukkala E. Nationwide registry-based analysis of cancer clustering detects strong familial occurrence of kaposi sarcoma. *PLoS One*, 2013, 8(1):e55209.

- II **Aavikko M**, Kaasinen E, Nieminen JK, Byun M, Donner I, Mancuso R, Ferrante P, Clerici M, Brambilla L, Turlaki A, Sarid R, Guttman-Yassky E, Taipale M, Morgunova E, Pekkonen P, Ojala PM, Pukkala E, Casanova JL, Vaarala O, Vahteristo P, Aaltonen LA. Whole genome sequencing identifies *STAT4* as a putative susceptibility gene in classic Kaposi sarcoma. *Manuscript*.

- III Saarinen S, **Aavikko M**, Aittomäki K, Launonen V, Lehtonen R, Franssila K, Lehtonen HJ, Kaasinen E, Broderick P, Tarkkanen J, Bain BJ, Bauduer F, Ünal A, Swerdlow AJ, Cooke R, Mäkinen MJ, Houlston R, Vahteristo P, Aaltonen LA. Exome sequencing reveals germline *NPAT* mutation as a candidate risk factor for Hodgkin lymphoma. *Blood*, 2011, 118(3):493-498.

- IV **Aavikko M^{*}**, Li SP^{*}, Saarinen S^{*}, Alhopuro P, Kaasinen E, Morgunova E, Li Y, Vesanen K, Smith MJ, Evans DG, Pöyhönen M, Kiuru A, Auvinen A, Aaltonen LA, Taipale J, Vahteristo P. Loss of *SUFU* function in familial multiple meningioma. *The American Journal of Human Genetics*, 2012, 91(3):520-6.

^{*} Equal contribution.

The publications are reproduced with the permission of the copyright holders. Study I is included in the doctoral thesis “*Large-scale data analysis to identify novel disease phenotypes and genes*” by Eevi Kaasinen (Helsinki, 2014) and study III in the doctoral thesis “*Genetic Basis of Familial Lymphoma Predisposition*” by Silva Saarinen (Helsinki, 2013).

ABBREVIATIONS

AASS	Amino adipate-semialdehyde synthase	KIT	v-kit Hardy-Zuckerman 4 feline sarcoma viral oncogene homolog
AIDS	Acquired immune deficiency syndrome	KS	Kaposi sarcoma
ALK	Anaplastic lymphoma receptor tyrosine kinase	LOD	Logarithm of odds
ATM	Ataxia telangiectasia mutated	LOH	Loss of heterozygosity
bp	base pair	MAF	Minor allele frequency
BRCA1, 2	Breast cancer 1, 2 early onset	MEF	Mouse embryonic fibroblast
BRIP1	BRCA1 interacting protein C-terminal helicase 1	MIM	Mendelian inheritance in man
BWA	Burrows-Wheeler aligner	MLH1	mutL homolog 1
CCR5	Chemokine (C-C motif) receptor 5 (gene/pseudogene)	MSH2, 6	mutS homolog 2, 6
CGH	Comparative genomic hybridization	MYC	v-myc avian myelocytomatosis viral oncogene homolog
cKS	classic Kaposi sarcoma	NF2	Neurofibromin 2 (merlin)
cM	centimorgan	NLPHL	Nodular lymphocyte predominant Hodgkin lymphoma
CNV	Copy number variant	NPAT	Nuclear protein, ataxia-telangiectasia locus
DNA	Deoxyribonucleic acid	NPR	National Population Registry
EBV	Epstein-Barr virus	PBS	Phosphate buffered saline
ENCODE	Encyclopedia of DNA elements	PBMC	Peripheral blood mononuclear cells
FFPE	Formalin-fixed, paraffin-embedded	PCR	Polymerase chain reaction
FCR	Finnish Cancer Registry	PFA	Paraformaldehyde
FCS	Fetal calf serum	PMA	Phorbol 12-myristate 13-acetate
FIMM	Institute for Molecular Medicine Finland	PTCH1, 2	Patched 1, 2
GATK	the Genome Analysis Toolkit	PTEN	Phosphatase and tensin homolog
GLI1, 2, 3	GLI family zinc finger 1, 2, 3	RB1	Retinoblastoma 1
GWAS	Genome-wide association study	RET	Ret proto-oncogene
HBV	Hepatitis b virus	RNA	Ribonucleic acid
HCV	Hepatitis c virus	SMARCE1	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily e, member 1
Hh	Hedgehog signaling	SNP	Single nucleotide polymorphism
HHV8	Human herpesvirus 8	SMO	Smoothed, frizzled class receptor
HIV	Human immunodeficiency virus	STAT4	Signal transducer and activator of transcription 4
HL	Hodgkin lymphoma	SUFU	Suppressor of fused homolog (Drosophila)
HLA	Human leukocyte antigen	SV	Structural variation
HPV	Human papilloma virus	TP53	Tumor protein p53
HTLV-1	Human T-lymphocytic virus 1	VHL	von Hippel-Lindau tumor suppressor, E3 ubiquitin protein ligase
IARC	International Agency for Research on Cancer		
ICGC	International Cancer Genome Consortium		
IFN- γ	Interferon gamma		

ABSTRACT

Tumors are complex diseases that are caused by a combination of environmental, lifestyle and genetic factors and they can originate from almost any cell type of the human body. Majority of tumors occur in isolated patients, but a fraction aggregate in families. In rare cases, familial aggregation of tumors is caused by inheritance of a mutated gene, which results in a high risk to develop specific tumors. About 1-5% of common cancers are estimated to be caused by inheritance of such genes. Although these patients constitute only a small fraction of all cancer patients, identification of these genes has improved significantly our understanding of the molecular basis of cancer as well as patient care and surveillance.

Over one hundred high risk tumor susceptibility genes have been identified during past few decades, but some remain yet to be characterized. These include those that do not cause other clinically recognizable syndromic features, and present with varying or incomplete penetrance. One aim of this study was to assess the familial aggregation of cancers in Finland and to identify novel tumor susceptibility conditions and families by systematically analyzing the population based patient data at the Finnish Cancer Registry (FCR). Another purpose of this study was to identify and characterize novel tumor susceptibility genes from families obtained from the FCR search as well as from medical genetics clinics.

In the first study, to assess the familial aggregation of tumors and to identify novel tumor susceptibility families, we conducted a systematic computerized clustering of the entire FCR patient database. Altogether 878 593 patients, diagnosed in 1952-2011, were clustered based on family name at birth, municipality of birth, and tumor type. Additional clustering based on family name at birth and tumor type was also conducted. For all clusters, observed and expected ratio, with 95% confidence intervals, were calculated. This yielded 25 910 family name at birth, municipality of birth, and tumor type based clusters, and 12 695 family name at birth and tumor type based clusters, representing 183 different tumor types.

We assessed the familial occurrence of the different tumors by a cluster score that proportions the patients in the most significant clusters to the frequency of the given tumor type in Finland. As expected, ranking of the tumor types based on the cluster score showed that most frequently clustered tumor types included those of well-known genetic background, such as thyroid medullary carcinoma, Wilms' tumor and retinoblastoma, but also those of unknown genetic etiology such as Kaposi Sarcoma (KS). We performed genealogy analysis on the clustered KS patients and showed that 70% of them were relatives, forming one family with five affected individuals in two generations, and several smaller families with two first degree relatives with KS. We also showed that KS incidence was higher in the individuals originating from western and northeastern Finland compared to individuals originating from other parts of the country.

In the second study, the genetic susceptibility of KS in the families that were identified in study I was examined. We mapped the shared chromosomal regions and performed whole genome and exome sequencing and identified 14 protein code altering candidate variants, among them a c.1337 C>T (p.Thr446Ile) variant in *signal transducer and activator of transcription 4*, (*STAT4*). The variant was not found from 242 Finnish genomes or 180 healthy regional control individuals.

We collected samples from the c.1337C>T variant carriers and showed that the activated T-cell populations of the carriers had decreased interferon gamma (IFN- γ) production. We studied the protein coding regions of *STAT4* and related genes in the exomes sequenced from seven Finnish familial KS cases and identified a missense mutation p.His289Arg in *chemokine (C-C motif) receptor 5 (gene/pseudogene) (CCR5)*, that segregated with the KS in the family with two affected siblings. We also studied the entire protein coding region of *STAT4* from 13 additional familial KS cases and the *STAT4* c.1337C>T site from 56 sporadic KS tumors, but did not find additional mutations. Although our results were not genetically validated, they suggest *STAT4* as a KS predisposing gene. However, further genetic and functional validation is needed to claim causality.

In study III, a Finnish family of four cousins with a rare subtype of Hodgkin lymphoma, nodular lymphocyte predominant Hodgkin lymphoma (NLPHL), was studied. To identify the genetic defect underlying the NLPHL susceptibility, we performed exome sequencing together with genome-wide linkage analysis and found a two base pair (bp) frameshift deletion (c.2437-2438_delAG) in *nuclear protein, ataxia-telangiectasia locus (NPAT)* segregating with NLPHL in the family. The patients had a decreased expression of *NPAT*, measured from the blood. We studied *NPAT* from a large number of other patients with Hodgkin lymphoma, and identified a three bp in-frame deletion c. 2171-2173delCTT (p.Ser725del), which was shown to be more prevalent in cases than population matched controls (odds ratio 4.11, P=0.018), supporting the role of *NPAT* as a possible predisposition gene for NLPHL.

A similar approach was used in study IV, where a Finnish family with five affected siblings with *NF2*-negative multiple intracranial meningiomas was scrutinized. By combining genome-wide linkage analysis and exome sequencing we identified a missense mutation c.367C>T (p.Arg123Cys) in *suppressor of fused homolog (Drosophila) (SUFU)* to segregate with the meningiomas in the family. We studied the tumors for loss of heterozygosity and found that all seven studied tumors had lost the wild-type allele, supporting the role of *SUFU* as a classic tumor-suppressor gene. We studied the effect of the mutation *in vitro* and *in silico* and showed that the mutated *SUFU* had lowered activity leading to dysregulated hedgehog signaling, most likely due to disrupted tertiary structure of the protein. We also studied the presence of additional *SUFU* mutations in a large set (n=162) of meningioma patients, but did not identify additional pathogenic mutations. The data indicates that, in addition to medulloblastomas, germline mutation in *SUFU* predisposes to meningiomas, particularly to multiple and intracranial types.

REVIEW OF THE LITERATURE

1. Tumorigenesis

Cell is the basic unit of all living organisms. The human body consists of about $3-4 \times 10^{13}$ cells (Bianconi *et al.*, 2013), representing hundreds of different cell types, that are all descendants of the fertilized egg. Most of the cells in the adult human body are carefully programmed to serve a defined function and only in collaboration with one another. For maintaining tissue and to carry out functions such as wound healing or menstrual cycle, cells need to retain the ability to replicate and renew throughout the lifetime of an individual. Although the replication machinery is deliberately controlled, it becomes time to time corrupted, which can lead to uncontrolled cell division and growth and eventually to a formation of an atypical mass of cells, a tumor.

Tumors can be divided into benign and malignant based on their ability to invade the tissues locally, and seed cells to other tissues and organs, that is, to metastasize. Compared to malignant tumors, benign tumors generally grow slower, resemble more the tissue they derive from and are rarely life threatening to their host. However, they can cause serious health problems by pressing the nearby tissues and anatomical structures, inducing for example neurological deficits and/or over-secretion of hormones.

A common concept in tumorigenesis is that malignant tumors develop from benign predecessors. This has been well-established especially in tumors that arise from epithelial cells, such as colorectal cancer (Fearon and Vogelstein, 1990). The great majority (~85%) of colorectal cancers are estimated to develop from a benign adenomatous polyp and progress into a mature colorectal cancer through several intermediate forms, in a process called the adenoma-carcinoma sequence (Fearon and Vogelstein, 1990; Vogelstein *et al.*, 1988). Some benign tumors, such as leiomyomas and meningiomas, however, rarely advance to malignant tumors (Lall *et al.*, 2013; van Meurs *et al.*, 2012).

The development of a tumor is a lengthy process that can take years or even decades, and it involves reprogramming of several aspects of the tumor forming cells. A landmark paper by Douglas Hanahan and Robert A. Weinberg (2000) described the essential biological features, ‘the hallmarks of cancer’ (I-VI, Box 1), that a cell has to obtain in order to initiate cancer. A decade later the authors complemented the list with two ‘emerging hallmarks’ (VII and VIII, Box 1) that are conceivably relevant to tumorigenesis (Hanahan and Weinberg, 2011). ‘Genome instability and mutation’ as well as ‘tumor promoting inflammation’ were suggested to act as facilitators for gaining the biological features needed for the tumor development (Hanahan and Weinberg, 2000; Hanahan and Weinberg, 2011). Of the eight ‘hallmarks of cancer’, all others are applicable to benign and malignant tumors except VI (the ability to invade nearby tissues and seed distant metastasis), which is applicable only to malignant tumors.

Box 1. The hallmarks of cancer (Hanahan and Weinberg, 2000; Hanahan and Weinberg, 2011)

I ‘Self-sufficiency in growth signals’

Growth inducing signals are required for normal cell proliferation. One capability that tumor cells have acquired is that they are no longer dependent on the outside growth signals. Dividing tumor cells either signal the neighboring cells to produce growth signals, learn to produce an excess of these themselves, or change their behavior as if they would be present all the time.

II ‘Insensitivity to anti-growth signals’

In addition to the growth inducing signals, the environment of a normal cell contains multiple growth inhibitory signals. To prosper, a tumor cell needs to acquire ways to overcome growth inhibitory signals.

III ‘Evading apoptosis’

When normal cell is under an unusual stress caused by, for example, excessive DNA damage or lack of nutrients and/or oxygen, the cell is directed to undergo programmed cell death, apoptosis. In order to grow, a tumor cell needs to acquire means to eliminate signals directing it to various cellular death pathways.

IV ‘Limitless replicative potential’

The number of cell divisions is by implication measured by the shortening of the telomeres. This is how the cell keeps track of its growth and the number of divisions it has gone through. Most of the normal cells can undergo only a limited number of cell divisions after which they stop dividing and reach irreversible quiescent state called senescence. In order to divide endlessly, tumor cell needs to learn to manipulate its time keeping machine, the telomeres.

V ‘Sustained angiogenesis’

To obtain enough oxygen and nutrition and to get rid of metabolic waste and carbon dioxide, cells have to be located within <1 mm from a capillary blood vessel. The same holds for tumor cells. Thus expanding tumor needs to be able to induce growth of new vasculature.

VI ‘Tissue invasion & metastasis’

The capability of a tumor cell to invade and metastasize requires abilities to escape the primary tumor mass as well as colonize new ground. Thus the process of invasion and metastasis is considered to require large scale reprogramming of many cellular functions, especially those that have to do with cell-cell adhesion, extracellular matrix degradation and motility.

VII ‘Reprogramming Energy Metabolism’

To support proliferation, tumor cells adjust their energy supply. This can be done by reprogramming glucose metabolism, increasing the number of glucose transporters on the cell surface or, relying on alternative metabolic pathways and/or energy sources. Many tumors change to glycolysis even in normal oxygen conditions. Despite the fact that, this is not the most cost-effective way to produce energy, it allows tumor cells to supply glycolytic intermediates to cellular pathways that are needed to build new cells.

VIII ‘Evading immune destruction’

The theory of immune surveillance suggests that the immune system constantly monitors the human body for abnormal cells. Based on this idea, arising tumors have somehow managed to avoid the detection and destruction by the immune system. The important role of immunological monitoring can be seen, for example, in the increased burden of tumors in immunocompromised individuals.

Two models are commonly proposed to explain tumorous growth (Beck and Blanpain, 2013; Weinberg, 2013). The stochastic model is based on the idea that each cell is equipotent to gain replicative advantage or differentiate. In this model, any cell clone with growth advantage forms a proliferating clonal population (Beck and Blanpain, 2013; Weinberg, 2013). The other model, the tumor stem cell model, is based on the idea that a higher organization exists in the tumors, much like in the normal tissues. In this model, only a small subset of cells, called the tumor stem cells, are able to proliferate and drive the tumor development (Reya *et al.*, 2001; Dick, 2008; Beck and Blanpain, 2013).

2. Human genome

With few exceptions, all human cells contain the same set of genetic material, the human genome. It is composed of deoxyribonucleic acid (DNA) and organized into 22 pairs of linear autosomal chromosomes, two linear sex chromosomes and a small circular mitochondrial genome. The linear chromosomes are deposited in the cell nuclei, and the mitochondrial genome is stored within the mitochondria, which reside in the cytoplasm.

The first drafts of the human genome sequence were published in 2001 (Lander *et al.*, 2001; Venter *et al.*, 2001), followed by the first complete version of the human genome sequence in 2004 (International Human Genome Sequencing Consortium, 2004). Since 2004, the human genome sequence has been corrected and improved, and today probably the most up-to-date human reference genome sequence is upheld and updated by Genome Reference Consortium. The human genome reference assembly, most in use currently, contains about 3.3 billion nucleotides, or DNA monomers (Genome Reference Consortium, Human Genome Assembly GRCh37, released in March 2009 <http://www.ncbi.nlm.nih.gov/projects/genome/assembly/grc/human/>, latest update GRCh37.p13, June 2013). According to the Encyclopedia of genes and gene variants (the GENCODE project), the reference human genome encodes 57 820 genes of which 20 345 translate proteins, 14 206 are pseudogenes, 13 870 transcribe long non-coding RNAs, 9013 transcribe small non-coding RNAs and 386 encode protein coding segments of the immunoglobulin and T-cell receptor gene segments (The GENCODE project, version 19 July 2013 freeze, GRCh37, <http://www.genencodegenes.org/stats.html>).

The coding parts of the human reference genome constitute less than 3% of its entire length (Alexander *et al.*, 2010) and the rest is composed of so called non-coding regions, which were historically considered as “junk DNA”, with little or no biological function. During the last few years, however, increasing numbers of non-coding sequence elements, relevant for example in the regulation of gene expression or DNA replication, have been annotated (Maston *et al.*, 2006; Harmston and Lenhard, 2013). The international Encyclopedia of DNA Elements (ENCODE) project recently reported a biological function for more than 80% of the human genome (ENCODE Project Consortium *et al.*, 2012). Although ENCODE’s data interpretation and definition of the biological function has faced a lot of criticism (Doolittle, 2013; Eddy, 2012; Graur *et al.*, 2013; Niu and Jiang, 2013), it is evident that, in addition to the protein coding genes, biologically relevant sequences are widely spread throughout the non-coding parts of the human genome.

Each individual human genome contains a significant amount of variation compared to each other or to the human reference genome. Several international large-scale projects, such as the International HapMap Project (International HapMap Consortium, 2003; International HapMap Consortium, 2005; International HapMap Consortium *et al.*, 2007) and more recently the 1000 Genomes Project (1000 Genomes Project Consortium *et al.*, 2012), have been carried out to elucidate the genetic variation in individual genomes and different populations. In 2012, the 1000 Genomes Project Consortium published results on low-coverage genomes, high-coverage exomes and dense single nucleotide polymorphism (SNP) genotypes of 1092 individuals representing 14 different populations, including 93 Finnish individuals. They reported altogether 38 million SNPs, 1.4 million short insertions or deletions, and over 14 000 larger deletions (1000 Genomes Project Consortium *et al.*, 2012). Furthermore, they described that on average each individual genome contains over 3 million SNPs (on average 1 SNP in every 1000 nucleotides) including 2500 protein coding variants and 150 protein function disrupting variants (1000 Genomes Project Consortium *et al.*, 2012). Most of these are shared by many individuals and multiple populations, but some are rare and present only in a certain population or an individual. Over half (53%) of the rare variants (minor allele frequency; MAF <0.5%) reported by the 1000 Genomes Project Consortium were present only in one population (1000 Genomes Project Consortium *et al.*, 2012).

In addition to SNPs small insertions and deletions, human genomes vary on a DNA copy number level (Feuk *et al.*, 2006). Copy number variants (CNVs) represent gains and losses of larger areas of the genomic DNA, ranging from one kilobase (kb) to several megabases in size. The curated database of genomic variants includes currently more than 2.5 million CNVs (the Database of Genomic Variants, <http://dgv.tcag.ca/dgv/app/home>, accessed 25th of April 2014). It has been estimated that each individual harbors on average 200 CNVs, majority of which are 1-10 kb in size (Kidd *et al.*, 2008, and the Database of Genomic Variants).

Individual human genomes vary also in their contents of sequential nucleotide repeats, called micro- and minisatellites. Although this type of variation is usually considered neutral, it is known to cause for example certain neurological diseases (Brouwer *et al.*, 2009). Almost half of the human genome is also occupied by transposons and transposon-like elements (Mills *et al.*, 2007). Transposons are in principle mobile elements that can produce variation in the human genome by copying and pasting DNA pieces from one place to another. The only transposons that are currently known to be active in the human genome are retrotransposable elements, including long interspersed elements, short interspersed elements, and SVA (SINE/VNTR/Alu) elements (Callinan and Batzer, 2006; Mills *et al.*, 2007), but their contribution to genetic variation is yet to be fully elucidated.

The earliest settlements in Finland dates back to some 10 000 years, but the Finnish population is generally considered to be initiated by two major migratory waves 4000 and 2000 years ago (Jakkula *et al.*, 2008). A third major migratory wave, inhabiting the geographically wide areas of the northern and eastern Finland, took place in the 16th century. During this period isolated villages were established by a small number of settlers generating multiple genetically distinct isolated subpopulations (Jakkula *et al.*, 2008). The early founder effects together with the more recent bottle necks have resulted in the decreased genetic diversity and increased number of rare

and loss-of-function variants of Finnish population (1000 Genomes Project Consortium *et al.*, 2012; Jakkula *et al.*, 2008; Lim *et al.*, 2014).

3. Tumor genome

The idea that tumors develop due to genetic defects dates back to the times of a pioneering German biologist Theodor Boveri (1862-1915). He studied dividing sea urchin eggs and noticed that the embryonic development was successful only when the genetic material was correctly divided between the two daughter cells. This led him to postulate that human tumors might similarly be caused by unequal segregation of the genetic material (Boveri, 1902). Boveri went on with his ideas, and published later a theoretical essay hypothesizing several aspects of tumorigenesis, that were only later proven experimentally, including clonal evolution, development of malignant tumors from benign predecessors, presence of growth promoting and inhibiting factors (genes) in the chromosomes (Boveri, 1914; Boveri, 2008).

Since the times of Boveri, the developments made in microscopy and molecular biology, especially the discovery of the DNA structure (Avery *et al.*, 1944; Watson and Crick, 1953a; Watson and Crick, 1953b) and the genetic code (Crick *et al.*, 1961), have led to our current understanding, that tumors are genetic diseases caused by genomic alterations, or mutations (Nowell, 1976; Hanahan and Weinberg, 2011), which can be caused by endogenous or exogenous carcinogens or be produced by erroneous DNA replication. Mutations that affect the tumor genomes range from subtle nucleotide level alterations, including substitutions and small insertions and deletions, to a chromosome level alterations, including translocations, inversions, deletions or amplifications. Additionally, whole chromosomes can be lost or gained (aneuploidy) and even the entire set of chromosomes can be multiplied (polyploidy).

Other central flaws affecting the tumor genomes include epigenetic alterations that do not change the actual DNA sequence, but may lead to altered gene expression and thus assist the developing tumor to acquire novel profitable traits. From the epigenetic modifications affecting tumors, DNA methylation and histone modifications are probably the best studied. Methylation is a simple covalent DNA modification that occurs generally at sites of CpG nucleotides. Both hypomethylation and hypermethylation, especially at the promoter regions of tumor associated genes, are commonly observed in the tumor genomes (Choi and Lee, 2013). Hypomethylation usually leads to an increase and hypermethylation to a decrease of gene expression (Choi and Lee, 2013). Histones are proteins that are used to pack the human genome. They are regulated by post-translational modifications, and when this is done incorrectly it can also lead to altered gene expression (Choi and Lee, 2013).

The overall number and types of mutations varies greatly among different tumor types and between tumors. The highest numbers of mutations, up to thousands, are seen in tumors with defective DNA repair machinery, including microsatellite unstable colorectal or gastric cancers (Gryfe and Gallinger, 2001) and colorectal cancers harboring mutations in the proofreading domains of DNA polymerases *POLE* and *POLD1* (Cancer Genome Atlas Network, 2012; Palles *et al.*, 2013). High numbers of mutations, on average 200 per tumor, are also observed in tumors

with a potent exogenous mutagen, such as tobacco smoke exposed lung cancers and UV-exposed melanomas (Pfeifer, 2010; Vogelstein *et al.*, 2013). In tumors originating from rapidly renewing tissues, such as those deriving from epithelium, the patient age has been shown to directly correlate with the number of mutations (Tomasetti *et al.*, 2013). Least mutations are observed in certain benign tumors, pediatric cancers, and leukemias (Mardis *et al.*, 2009; Francis *et al.*, 2013; Mehine *et al.*, 2013).

The recent application of high-throughput sequencing methods has revolutionized the field of tumor genomics enabling very detailed and precise characterization of the mutational profiles of different tumors. In addition to the sequencing studies conducted in separate research laboratories and centers, several large-scale international projects have been put together to elucidate the mutational landscapes of many human tumors. Probably the largest of these is the International Cancer Genome Consortium (ICGC) that aims to catalogue the mutations of the 50 most common tumor types and release the data for public use (<https://www.icgc.org/> and International Cancer Genome Consortium *et al.*, 2010). Currently the ICGC database includes genomic sequences from 42 projects representing 18 different tumor primary sites from over 10 000 individuals (ICGC Data release 15.1, February 12, 2014, <http://dcc.icgc.org/>).

The novel sequencing methods have also provided a more detailed description of the mutational processes present in the tumors. In a recent study, 21 distinct ‘mutational signatures’ were identified from the genomic and exomic sequences of 7042 primary tumors deriving from 30 different tumor types (Alexandrov *et al.*, 2013). In addition to the already established mutational signatures, like those induced by carcinogens in tobacco, UV-light, and defective DNA repair machinery, novel signatures associated with, for example, patient age or usage of certain anticancer drugs were described. A few signatures remain yet to be explained (Alexandrov *et al.*, 2013) and some possibly yet to found.

Novel mutational phenomena have also been recently described. Chromotripsis, which was initially found as a frequent event in the genomes of malignant bone tumors (Stephens *et al.*, 2011), is a complex genomic rearrangement caused by chaotic shattering of one or few chromosome arms into tens to hundreds of pieces followed by a stochastic reassembly (Stephens *et al.*, 2011). Chromotripsis has been associated with advanced stage and poor prognosis (Forment *et al.*, 2012), but similar chromosomal alterations have later been shown in benign tumors (Mehine *et al.*, 2013; Parker *et al.*, 2014). Chromotripsis has also been reported to affect germline genomes and it was recently identified as the underlying mechanistic cause of specific Robertsonian translocation rob(15;21)(q10;q10)c, which is known to predispose to childhood acute lymphoblastic leukemia (Li *et al.*, 2014).

Kataegis, which is defined by a regional hypermutation of long stretches of DNA (kilobases in length), was originally identified from a subset of breast cancer genomes (Nik-Zainal *et al.*, 2012). It has also been subsequently found from ovarian cancer and multiple myeloma genomes (Bolli *et al.*, 2014; Hoogstraat *et al.*, 2014). The mutations in kataegis commonly affect the same parental chromosome, are preferably C>T transitions, and co-localize with chromosomal rearrangements. Although the mechanism causing Kataegis is not yet well understood, APOBEC family of

cytidine deaminases have been suggested to play a role (Lada *et al.*, 2012; Nik-Zainal *et al.*, 2012).

3.1 Driver mutations and genes

Genetic and epigenetic changes that affect the tumor genomes can be classified as ‘drivers’ or ‘passengers’ (Thiagalingam *et al.*, 1996). Driver mutations are mutations or epigenetic changes that confer a selective growth advantage to the cell whereas passenger mutations are bystanders and reflect the stochastic mutational process present in the tumors (Thiagalingam *et al.*, 1996). Tumors have been estimated to contain one to eight driver mutations, each of which most likely conferring only a small selective growth advantage (Armitage and Doll, 2004; Bozic *et al.*, 2010; Welch *et al.*, 2012; Vogelstein *et al.*, 2013). The rest of the mutations are passengers.

Driver mutations occur in a fraction of genes that can be called as ‘driver genes’. These are normal genes that in mutated form can promote tumorigenesis (Vogelstein *et al.*, 2013). Driver genes can further be classified into proto-oncogenes and tumor suppressor genes. Proto-oncogenes are normal cellular genes that encode proteins involved in the cell growth, cell replication, inhibition of cell differentiation, and regulation of cell death (Croce, 2008). Proto-oncogenes are activated by overexpression or gain-of-function mutations after which they are called oncogenes. At the cellular level, oncogenes are dominant, which means that usually only one of the two gene copies needs to be mutated to stimulate tumorigenic effects. Common oncogenic mutations are simple base substitutions changing a critical amino acid into another during protein translation, but oncogenes can also become activated by a translocation or amplification that leads to overexpression of the proto-oncogene. Well-established proto-oncogenes encode for example growth factors (such as *EGF* and *PDGFs*) and growth factor receptors (*EGFRs* and *PDGFRs*), chromatin re-modelers (*KMT2A*), signal transducers (*BRAF* and *KRAS*), transcription factors (*MYC*), and apoptosis regulators (*BCL2*).

Tumor suppressor genes normally function in growth repression, by either controlling cell proliferation or promoting apoptosis (Sherr, 2004). Mutations responsible for the tumorigenic actions of tumor suppressor genes are loss-of-function mutations and typically both copies of the tumor suppressor gene need to be affected for the tumorigenic effect. Mutations that commonly affect tumor suppressor genes are small insertions and deletions breaking the reading, nucleotide substitutions that cause a translation of a premature stop-codon, or mutations affecting mRNA splicing. However, they can also be larger chromosomal deletions that delete the tumor suppressor gene partially or entirely. Mutations found in tumor suppressor genes are usually not clustered and are spread across the length of the gene (Vogelstein *et al.*, 2013).

Tumor suppressor genes can be classified into gatekeepers, caretakers, (Kinzler and Vogelstein, 1997), and landscapers (Kinzler and Vogelstein, 1998). Gatekeepers are the most classical types of tumor suppressor genes, and their normal functions include inhibition of cell growth and proliferation. Many gatekeeper genes have been identified as the underlying causal genes behind well-established cancer susceptibility syndromes, such as Neurofibromatosis, type I (MIM 16220), as well as Li-Fraumeni (MIM 151623) and Von Hippel-Lindau (MIM 19330) syndromes. Caretakers on the other hand are stability genes that are responsible, for example, for maintaining

DNA intact and mitotic recombination. An inactivated caretaker has an indirect effect on tumorigenesis and it can, for example, increase the general mutation rate of the genome and thereby increase the chances of a mutation in another driver gene. Examples of caretakers are *MLH1* and *MSH2*, which encoding proteins central in the DNA mismatch repair machinery, and *BRCA1* and *BRCA2*, which encode proteins that repair DNA double strand breaks. Landscapers, such as *SMAD4*, remodel the tumor surroundings more amenable for growth.

Concepts of dominant negativity and haploinsufficiency are tightly linked to tumor suppressor genes (Santarosa and Ashworth, 2004; Payne and Kemp, 2005). Dominant negative mutations affect especially polymeric molecules, such as *TP53*, *ATM*, and *CDKN1A* (Payne and Kemp, 2005). The mutated protein product is not fully functional, but it is functional enough to interact with the wild-type protein product, thus impairing the actions of the entire polymeric protein. Haploinsufficiency means that inactivation of only one copy of a tumor suppressor gene has already a positive effect to tumor growth (Santarosa and Ashworth, 2004). Some studies have suggested that tumor suppressor genes and proto-oncogenes cluster in certain areas of the genome and that larger hemizygous chromosomal alterations that either delete several tumor suppressor genes or amplify several proto-oncogenes may introduce a selective growth advantage for the cell (Solimini *et al.*, 2012; Xue *et al.*, 2012).

4. Inherited susceptibility to tumors

Although most tumors arise in isolated patients, familial aggregation of tumors has been long acknowledged. One of the earliest reports of familial clustering of tumors was published by Paul Broca in 1866, who described his wife's family with 15 members with breast cancer (Broca, 1866). Majority of common cancers have been shown to display high concordance in monozygotic twins, compared to dizygotic twins or siblings, indicating importance of the shared genetic factors (Lichtenstein *et al.*, 2000). The relevance of the shared genetic factors has also been emphasized by a number of family- and population-based studies. Large population-based studies have been conducted, for example, using Utah Population and Cancer Registry Database, the Swedish Family-Cancer Database, or Icelandic Cancer Registry and the genealogic database. These studies have collectively demonstrated that most of the common tumor types display increased risk in first-degree relatives and also beyond the nuclear family (Cannon-Albright *et al.*, 1994; Goldgar *et al.*, 1994; Vaittinen and Hemminki, 1999; Dong and Hemminki, 2001; Czene *et al.*, 2002; Amundadottir *et al.*, 2004; Albright *et al.*, 2012). In addition to familial aggregation of specific cancer types within families, marked coaggregation of different cancers has also been reported (Thomas *et al.*, 1999).

In addition to the genetic factors, there are also a variety of environmental factors that may be shared within families, and a fraction of familial cancers is likely to be explained by these (Hemminki *et al.*, 2006). For example, increased risk for lung and stomach cancers has been reported in spouses of lung and stomach cancer patients, emphasizing the role of shared environment (Hemminki *et al.*, 2001; Hemminki and Jiang, 2002; Amundadottir *et al.*, 2004). Also certain childhood cancers, such as acute lymphoblastic leukemia, which show high concordance in monozygotic twins, have been shown to have non-constitutional, prenatal causes

(Ford *et al.*, 1993; Wiemels *et al.*, 1999). Nonetheless, even in these cases the individual genetic make-up most likely plays a central role, since it influences the ability of an individual to deal with the environmental exposures.

Inherited genetic factors that predispose to tumors can be roughly divided into three categories: rare high risk variants, rare moderate risk variants and common low risk variants. The rare high risk variants are present in the population with minor allele frequency (MAF) lower than 0.1%, and the carriers of these are highly likely to develop a tumor (odds ratio ≥ 10.0). The rare moderate risk variants are present in less than 2% of the population and individuals carrying these variants are at >2 fold risk of developing a given tumor. Common low-risk alleles are frequent, present in $>10\%$ of the population, and confer a small risk for the individual (odds ratio <1.5) (Fletcher and Houlston, 2010).

4.1 Rare high risk variants

In general it is considered that 5-10% of tumors arise in individuals with hereditary cancer syndromes, caused by rare high risk variants in tumor predisposition genes (Nagy *et al.*, 2004). Examples of well-characterized hereditary cancer syndromes are, for example, hereditary breast and ovarian cancer syndrome (MIM #604370), Lynch syndrome (MIM #120435), Von Hippel Lindau syndrome (MIM #193300) and Li Fraumeni syndrome (MIM #151623). Clinical evidence for the presence of hereditary cancer syndrome include young age at onset, presence of bilateral disease, multiple primary tumors, and a family history with multiple family members affected by the same tumor type. Hereditary cancer syndromes are also often characterized by other non-tumor related clinical manifestations. These may be even more frequent in the family than the tumors, and they are often very important cues for the clinical diagnosis (Rahman, 2014). The most common non-tumor related clinical manifestations are various skin lesions. Also many neurological, dysmorphic, and skeletal features, such as microcephaly, macrocephaly, short stature, and developmental delay, are encountered in the carriers of the high risk alleles (Rahman, 2014).

In the beginning of the 20th century, multiple large families with clear inheritance of hereditary cancer susceptibility syndromes were described by clinicians such as Aldred S. Warthin (Warthin, 1929), Johannes Peutz (Peutz, 1921), and Harold Jeghers (Jeghers *et al.*, 1949), but it was not until 1987, when the first gene, *RBI*, was identified as the underlying gene in childhood retinoblastoma predisposition (Fung *et al.*, 1987). This finding was facilitated by the ‘two-hit hypothesis’ proposed by Alfred Knudson in 1971 (Knudson, 1971). Knudson studied the kinetics of familial and sporadic childhood retinoblastoma and noticed that the familial cases usually had earlier age at onset and their tumors affected more often both eyes. He proposed that retinoblastoma was caused by a single gene defect and that the development of retinoblastoma required both alleles of the gene to be inactivated. He suggested that in the familial cases the other allele of the gene is inherited as defective and the other is inactivated by somatic mutation later during life. In the sporadic cases, both alleles needed to be inactivated by *de novo* mutations (Knudson, 1971). After the discovery of the *RBI*, more than 100 rare high risk tumor predisposition genes have been identified (Marsh and Zori, 2002; Genuardi, 2004; Rahman, 2014).

The high risk tumor predisposition genes are most often tumor suppressor genes and follow Knudson's 'two-hit hypothesis'. Majority of them are located in autosomes and inherited dominantly, although recessive, X-linked and Y-linked forms also exist (Rahman, 2014). In addition, 16 high risk tumor predisposing genes have been described to display both dominant and recessive inheritance. Some of these, such as *BRCA2*, *MLH1*, *MSH2*, and *MHS6* have been associated with adult onset cancers when inherited dominantly and childhood cancer when inherited recessively (Rahman, 2014). A small minority of the high risk cancer susceptibility genes harbor also oncogenic gain-of-function mutations (Rahman, 2014). Examples of such include *RET*, *KIT*, and *ALK* (Rahman, 2014). Clinical symptoms are present in majority of the patients carrying high risk tumor predisposing alleles by the age of 70 (Nagy *et al.*, 2004). However, due to factors such as phenotypic variability and age related penetrance, some individuals carrying high risk tumor predisposing alleles do not show complete penetrance.

4.2 Rare moderate risk variants

A proportion of familial aggregation of common cancers has been detected to result from rare variants with incomplete penetrance conferring moderate risk to their carriers. These variants have been studied especially in breast cancer, where relatively rare alleles of different members of the DNA-damage response pathway, namely *CHEK2* (Meijers-Heijboer *et al.*, 2002; Vahteristo *et al.*, 2002), *PALB2* (Rahman *et al.*, 2007), *ATM*, (Renwick *et al.*, 2006), *NBS1* (Steffen *et al.*, 2006), and *BRIP1* (Seal *et al.*, 2006), have been identified to increase the risk of breast cancer 2-9 fold (Hollestelle *et al.*, 2010). In addition to breast cancer, inactivating mutations in *BRIP1* have been reported to confer moderate risk for ovarian cancer (Rafnar *et al.*, 2011), and a rare missense mutation in *HOXB13* has been reported to confer a moderate risk for prostate cancer (Xu *et al.*, 2013). A few examples exist, where a specific mutation in a well-established high risk cancer predisposition gene, has been associated with moderate risk. One example is the p.Glu1317Gln missense mutation in *APC* found in the Ashkenazi Jews (Laken *et al.*, 1997; Liang *et al.*, 2013). The presence of the rare moderate risk variants varies in populations, which makes it hard to validate the results and evaluate the actual risk (Lalloo and Evans, 2012).

4.3 Common low risk variants

The common disease-common variant hypothesis predicts that genetic variants that are frequent in the population affect the individual risk of getting a common disease (Risch and Merikangas, 1996). Common variants associated with common diseases can be studied by association studies that compare the frequencies of alleles in sets of unrelated cases and controls. Before development of the genome-wide genotyping methods, association studies were conducted using only a restricted number of common polymorphisms in selected candidate genes or pathways (Easton and Eeles, 2008; Chung and Chanock, 2011). Today, the candidate gene based methods have been largely replaced with genome-wide association studies (GWAS) (Hirschhorn and Daly, 2005), which measure the association of the disease with usually hundreds of thousands of markers distributed throughout the genome.

To date, about 400 different cancer-associated common low risk variants, or loci, have been identified through GWAS studies (Hindorff *et al.*, 2009; Rahman, 2014). The overall risk of an individual associating locus or SNP is generally small, in the order of 1.1-1.4 fold (Chung and Chanock, 2011). However, the combinatorial effect of multiple common low risk alleles may be substantially larger (Zheng *et al.*, 2008).

A major limitation with the GWAS is that the associated genomic markers are only landmarks of the associated loci, and many times the factors that actually cause the predisposition are difficult to identify. The underlying causal mechanisms have been determined for only a very small number of associated loci (Edwards *et al.*, 2013). Majority of the common low risk loci are associated with increased risk for one specific tumor type, but a few multi-cancer loci exist that are associated with multiple tumors, including the region flanking the *MYC* oncogene at 8q24, the *TERT-CLPTMIL* containing region at 5p15.33, and the ~400kb region at 11q13 (Chung and Chanock, 2011). Surprisingly few are located within high or moderate risk cancer genes, suggesting that the mechanisms underlying the predisposition caused by rare high and moderate risk alleles and common low risk alleles are in large part different (Rahman, 2014).

5. Infectious causes of cancer

Approximately 16-20% of cancers are estimated to be induced by infectious agents (Bouvard *et al.*, 2009; de Martel *et al.*, 2012). The contribution of the infectious agents to the entire cancer burden varies widely in different areas of the world, the biggest burden being in the low-resource countries where the infectious conditions are more widespread (de Martel *et al.*, 2012; IARC Working group, 2012a).

International Agency for Research on Cancer (IARC) Monographs programme has classified one bacterial (*Helicobacter pylori* [*H pylori*]), seven viral (Epstein-Barr virus [EBV], human T-cell lymphotropic virus type 1 [HTLV-1], human papilloma virus [HPV], human herpesvirus 8 [HHV8], hepatitis B and C viruses [HBV and HCV], and human immunodeficiency virus-1 [HIV-1]) and three parasitic infections (*Opisthorchis viverrini*, *Clonorchis sinensis*, and *Schistosoma haematobium*) as the Group 1 Biological carcinogenic agents to humans (Table 1). These agents are estimated to cause 2 million new cancers per year and more than 90% of these are caused by the top four agents, namely HBV and HCV, HPV and *H pylori*, (de Martel *et al.*, 2012, Table 1).

Mechanisms of how infectious agents can promote tumorigenesis include induction of chronic inflammation and reprogramming of the host inflammatory and infected cells to produce tumor growth enhancing molecules, which can further result to the production of carcinogenic agents (for example reactive oxygen species) and/or changes in the tumor microenvironment (Table 2). Four viral agents, HTLV-1, EBV, HHV8, and HPV, have also been established to have direct carcinogenic effects, meaning that they can *in vitro* immortalize cultured cells and express multiple oncogenes that interact with the host proteins, which in turn lead to alterations in the cell cycle control (Table 2). The genome of these agents is entirely or partly present in every cell in the tumor. Infectious carcinogenic agents can also cause immunosuppression. The best example of this is possibly HIV-1 infection which predisposes the affected individuals to opportunistic infections, including those with carcinogenic effects (Table 2).

Table 1. Group 1 biological carcinogenic agents assessed by the IARC Monograph Working Group

<i>Infectious agent</i>	<i>Associated cancers*</i>	<i>Infection attributable cancers per year**</i>
Helicobacter pylori	Gastric cancer, low-grade B-cell mucosa-associated lymphoid tissue (MALT) gastric lymphoma	660 000 (32.5%)
Human papillomavirus (HPV)	Carcinoma of cervix, vulva, vagina, penis, anus, oral cavity, oropharynx and tonsil	610 000 (30.0%)
Hepatitis B and C virus (HBV and HBC)	Hepatocellular carcinoma, non-Hodgkin lymphoma; HBC	600 000 (29.5%)
Ebstein-Barr virus (EBV)	Burkitt lymphoma, nasopharyngeal carcinoma, immune-suppression-related non-Hodgkin lymphoma, Hodgkin lymphoma, T-cell and NK-cell lymphoma (nasal type)	110 000 (5.4%)
Human Herpesvirus 8 (HHV8)	Kaposi sarcoma, primary effusion lymphoma	43 000 (2.1%)
<i>Schistosoma haematobium</i>	Urinary bladder cancer	6000 (0.3%)
Human T-cell lymphotropic virus type 1 (HTLV-1)	Adult T-cell lymphoblastic leukemia and lymphoma	2100 (0.1%)
<i>Opisthorchis viverrini</i> and <i>Clonorchis sinensis</i>	Cholangiocarcinoma, hepatocellular carcinoma	2000 (0.1%)
Human Immunodeficiency virus-1 (HIV-1)	Viral cancers, including: Kaposi sarcoma; KSHV, non-Hodgkin lymphoma; EBV related and non-related, Hodgkin lymphoma; EBV, Carcinoma of the cervix, anus, conjunctiva; HPV	Not determined
Total		2 000 0000 (100%)

*For which there is sufficient evidence in humans (IARC Working group, 2012a); **Numbers describes those of 2008 and are obtained from de Martel *et al.*, 2012.

Table 2. Properties of the human carcinogenic viruses

<i>Virus</i>	<i>Genome</i>	<i>Host genome integration</i>	<i>Infection related symptoms</i>	<i>Mechanistic effects</i>
EBV	DNA	no	Asymptomatic, respiratory illnesses and mononucleosis	Cell proliferation, inhibition of apoptosis, genomic instability, cell migration
HPV	DNA	yes	Asymptomatic, genital warts	Cell immortalization, genomic instability, inhibition of DNA damage response, anti-apoptotic activity
HBV	DNA	yes	Asymptomatic, acute hepatitis B	Inflammation, liver cirrhosis, liver fibrosis
HCV	RNA	nd	Asymptomatic	Immortalization and transformation of T cells
HTLV-1	RNA	yes	Asymptomatic, Myelopathy/ Tropical Spastic Paraparesis, inflammation of eye, joints, muscles, lung and skin	Cell proliferation, inhibition of apoptosis, genomic instability, cell migration
HHV8	DNA	no	Asymptomatic	Cell proliferation, inhibition of apoptosis, genomic instability, cell migration
HIV-1	RNA	yes	Immunosuppression	Immunosuppression (indirect effect)

6. Identification of rare high risk tumor susceptibility genes

Majority of the currently known rare high risk tumor susceptibility genes have been discovered by positional cloning, which means that the gene is identified based on its genomic location, without prior hypothesis of its biological function. The strategy that has been used the most in the identification of rare high risk tumor susceptibility genes is genome-wide linkage analysis, which is a family based statistical method to trace disease-associated genomic markers (Rahman, 2014). In addition, some tumor susceptibility genes have been successfully identified through candidate-gene based approaches. Screening of candidate genes based on their functional similarity to known tumor susceptibility genes has been used to identify novel susceptibility genes, such as *PMS6*, *MSH6*, *SDHD*, and *SDH5* (Nicolaidis *et al.*, 1994; Miyaki *et al.*, 1997; Niemann and Muller, 2000; Hao *et al.*, 2009). Another advantageous candidate-based approach has been to explore genes associated with distinct cellular susceptibility phenotypes, such as DNA repair defects (Nichols *et al.*, 1996; Sijbers *et al.*, 1996) or mosaicism (Hanks *et al.*, 2004). Also a small number of tumor susceptibility genes, including *TP53*, *KIT*, and *SUFU* have been identified because they are frequently somatically mutated in cancers (Malkin *et al.*, 1990; Nishida *et al.*, 1998; Taylor *et al.*, 2002). More recently, genome-wide mutation analyses using high-throughput sequencing either alone or in combination with linkage analysis have proven very powerful in the discovery of novel high risk tumor susceptibility genes (Comino-Mendez *et al.*, 2011; Testa *et al.*, 2011; Smith *et al.*, 2013).

6.1 Linkage analysis

Linkage analysis is based on the concept that chromosomal loci that are physically close to each other are linked and are usually inherited together and are only rarely separated by recombination during meiosis. On the contrary, when two loci are far apart, for example in different chromosomes or in the opposite ends of the chromosome, recombination is likely to happen and the two loci are inherited independently (Pulst, 1999). Meiotic recombinations are not evenly spread across the human genome, and they tend to cluster in so called recombination hot spots (Kauppi *et al.*, 2004; Paigen and Petkov, 2010). The observed portion of recombinations, that is recombination fraction (θ), is used to measure the relative distance between two loci in centiMorgans (cMs): one cM corresponds to a recombination fraction of 0.01 (1%). If θ is equal to 0.5, the two loci show no linkage and they are inherited independently. Therefore, if the two loci are linked, the θ is smaller than 0.5 (Pulst, 1999). The background frequency of recombinations is therefore determined by experiential genetic maps that describe a relative distance of loci in cMs; the greater the recombination frequency, the longer the distance between the two loci (Pulst, 1999).

In family-based linkage analysis, polymorphic markers are genotyped from affected and healthy members of the family, which are then used to test for co-inheritance with the disease. Based on the inheritance model of the disease in the family, the affected family members are expected to share alleles, at the disease causing genomic region, either in one chromosome (dominant and X-linked inheritance) or in both sister chromosomes (recessive inheritance).

In parametric linkage analysis, the genetic model for the disease inheritance is specified, and the penetrance as well as the allele frequency of the disease allele is defined. Nonparametric linkage analysis is a model-free method, in which no assumption of the inheritance model needs to be made. Linkage analysis, where multiple markers (more than two) are used to evaluate the linkage of the markers with the disease, is called multipoint linkage analysis. This increases the informativeness of a series of markers and enables effective detection of the linked chromosomal regions. In multipoint linkage analysis, the hypothesis of linkage between the disease gene and a set of markers is compared to the hypothesis of no linkage between the disease gene and a set of markers (Kruglyak *et al.*, 1996).

The statistical significance of linkage is estimated by the logarithm of odds (LOD) score (Morton, 1955). The LOD score is a 10^{th} base logarithm of the likelihood that the two loci are linked compared to the likelihood that the loci are not linked. Positive LOD values indicate that linkage exists and negative that there is no linkage. By convention, a LOD score over 3 denotes a strong likelihood that the genetic variant underlying the trait is located close to the marker. In practice, such LOD score is very hard to obtain from a single family, but the test allows summing of the LOD scores from different families.

6.2 High-throughput sequencing

Since the completion of the first human genome sequence (Lander *et al.*, 2001; Venter *et al.*, 2001; International Human Genome Sequencing Consortium, 2004), major improvements in sequencing technologies have been made. Developments especially concerning sample preparation and automation as well as parallelization of the sequencing reactions have led to significant increase in the speed with concurrent reduction in costs (Mardis, 2008).

Today, various high-throughput sequencing, also known as next-generation or massive-parallel sequencing methods are available. The commonly used sequencing methods include Illumina (Illumina), the SOLiD System (Applied Biosystems), 454 Life Sciences (Roche), and Complete Genomics (BGI-Shenzhen) (Mardis, 2008; Mardis, 2013). Using slightly different approaches, these technologies amplify single strands of DNA from a fragmented DNA library (usually 350-400bp in length) and sequence the amplified product. The libraries are produced by attaching a method specific linker and/or adapter sequences to the ends of the DNA fragments. These linker sequences enable the selective amplification of the fragmented DNA library in a massively parallelized PCR reaction and also the subsequent sequencing of the amplified products, or sequence reads (Mardis, 2008; Mardis, 2013).

In Illumina sequencing, DNA is first sheared to blunt ended fragments and adapter sequences are ligated to both ends. These enable the fragments to bind to the surface of a glass chamber called flow cell. When the fragments are attached to the flow cell surface, bridge amplification is performed to create a cluster of fragments consisting of up to 1000 copies of each fragment. For sequencing, Illumina uses sequencing by synthesis technology which uses four reversible terminated fluorescently labelled nucleotides. In each sequencing cycle, a single fluorescently labelled nucleotide is incorporated to the growing DNA sequence by DNA polymerase. The fluorescent signal is emitted and imaged to identify the incorporated nucleotide after which it is

enzymatically cleaved to enable addition of the next nucleotide in the growing DNA chain (Ulahannan *et al.*, 2013).

In most of the high-throughput sequencing studies employing human DNA, the sequencing reads are mapped against the human reference genome. A commonly used, relatively low memory using alignment method is the Burrows-Wheeler transformation (Li and Durbin, 2010), which is based on effective sorting and organization of strings of characters. An example of a widely used read mapping program is Burrows-Wheeler Alignment (BWA) (Li and Durbin, 2009; Li and Durbin, 2010). After read mapping, the nucleotides differing from the reference genome are detected using variant calling programs. Multiple statistical programs have been developed and commonly used programs include for example Sequence Alignment/Map tools (SAMtools, Li *et al.*, 2009), and Genome Analysis Toolkit (GATK, McKenna *et al.*, 2010).

Due to the massive amount of data that the high-throughput sequencing produces, the processing of the data is computationally demanding, and development of more efficient methods for data storage and handling is constant. Novel computational tools are also needed for the mapping, variant calling and analyzing the data. Especially difficult regions for the current sequencing and mapping technologies are repeated sequences, insertions and deletions and larger chromosomal alterations (Ulahannan *et al.*, 2013). The above mentioned difficulties can at least to some extent, be surpassed by novel single-molecule sequencing methods, such as the GridION™ (Oxford Nanopore Technologies) and the PacBio RS II (Pacific Biosciences), that do not necessary require amplification of the DNA prior sequencing and can sequence longer reads that are easier to uniquely map to the reference human genome.

7. Tumor susceptibility conditions scrutinized in this study

7.1 Kaposi sarcoma

Kaposi sarcoma (KS) is a malignant soft tissue sarcoma first described by Moritz Kaposi in 1872 as an indolent sarcoma of the skin affecting mainly elderly Ashkenazi Jewish men (Kaposi, 1872). KS commonly manifests as a slowly progressing single or multifocal purple lesions on the skin of the extremities or oral mucosa. However, the disease course is variable and KS can also manifest as an aggressive metastatic disease involving multiple organs.

Over hundred years later from the initial description of KS, clinicians in the major cities in the USA were faced with a sudden outbreak of KS among young homosexual men with acquired immunodeficiency syndrome (AIDS) (Friedman-Kien, 1981). Epidemiological evidence suggested involvement of an infectious agent underlying KS (Beral *et al.*, 1990; Lifson *et al.*, 1990), and in 1994 DNA sequences of a novel gamma herpesvirus, called Kaposi sarcoma-associated herpesvirus (KSHV or later named HHV8), were identified from the KS biopsies of AIDS patients (Chang *et al.*, 1994).

HHV8 is classified as Group 1 carcinogen to humans by IARC (IARC Working group, 2012b; Lifson *et al.*, 1990, Tables 1 and 2), and it is present in all KS tumors. The neoplastic cells are called ‘spindle cells’, and they are the main repository of HHV8 in the tumor (Boshoff *et al.*, 1995). In addition to the spindle cells, KS tumors are comprised of large infiltrates of normal cell

types including macrophages, lymphocytes, plasma cells and red blood cells (Lifson et al., 1990; IARC Working group, 2012b).

KS can be classified into four distinct epidemiological entities. These are classic KS (cKS) that affects mainly elderly men of Mediterranean or Eastern European Jewish origin; endemic KS, that involves individuals from Central and Eastern Africa; iatrogenic KS, affecting individuals that are on immunosuppressive medication; and epidemic or AIDS-related KS seen in HIV-infected individuals. The prevalence of KS varies geographically being extremely rare with age standardized incidence of less than 1 per 100 000 in males in most of the other continents whereas in the central parts of Africa it is the most common cancer with an incidence of over 22 per 100 000 in males (Parkin et al., 1999; Mesri et al., 2010). In Finland the overall incidence of KS is higher among males than in females, with an average incidence of 0.2 in males and 0.1 in females, per 100 000 person-years, age-adjusted to the World Standard Population (www.syoparekisteri.fi, updated 24.04.2014). In 2011, 14 new KS cases were diagnosed in Finland, of which 10 were males and 4 females (www.syoparekisteri.fi, updated 24.04.2014). In Nordic countries the incidence of KS has been reported to differ significantly, with Sweden having the highest and Denmark the lowest rates (Hjalgrim *et al.*, 1996).

Although HHV8 can be detected in all KS tumors, the infection alone is not sufficient to induce the development of KS. Genetic predisposition to cKS has been studied in a few unrelated children with recessive loss of function mutations, two with inherited immunodeficiency caused by mutations in *IFN- γ RI* and *WAS* (Camcioglu et al., 2004; Picard et al., 2006) and two idiopathic cases with mutations in *STIM1* and *TNFRSF4* (Byun et al., 2010; Byun et al., 2013). In addition, genetic variants in *FCGR3A*, *IL8RB*, and *IL-13* (Brown et al., 2005; Brown et al., 2006) as well as certain HLA-alleles have been associated with cKS predisposition, but these results have remained inconclusive.

7.2 Nodular lymphocyte predominant Hodgkin lymphoma

Hodgkin lymphoma (HL) is a group of malignant diseases of adaptive immunity. HL originates from lymphocytes, more precisely mature B-cells, and accounts for under 10% of all lymphoid malignancies (Swerdlow *et al.*, 2008). Hodgkin lymphoma involves two separate disease entities: classical HL, and Nodular lymphocyte-predominant HL (NLPHL) (Swerdlow *et al.*, 2008). Classical HL can be further subdivided into nodular sclerosing, mixed cellular, lymphocyte-rich and lymphocyte-depleted HL, of which most common is the nodular sclerosing type. Classical HL is a much more common HL subtype than NLPHL, which accounts only for 3-5% of the cases (Colby *et al.*, 1982; Morton *et al.*, 2006; Nogova *et al.*, 2008).

In Finland, 100-120 new HL cases are diagnosed every year, and the portion of NLPHLs is described to be as high as 16% (Saarinen *et al.*, 2013). NLPHL is usually diagnosed as a localized early stage disease and most common symptoms include lymphadenopathy of peripheral lymph nodes in the neck, groin, arm pit, or cervix (Swerdlow *et al.*, 2008; Lee and LaCasce, 2009). Compared to the classical HL, NLPHL has a slightly elevated age at diagnosis (30-50 years), unimodal age distribution, high male to female (3:1) predominance, and less aggressive clinical course with better prognosis (de Jong *et al.*, 2006; Nogova *et al.*, 2008).

Although NLPHL is rarely fatal, and its 10 year overall survival rate is over 80%, relapses happen frequently and a transformation to a non-HL, specifically to a diffuse large B-cell lymphoma, is sometimes observed (Al-Mansour *et al.*, 2010; Biasoli *et al.*, 2010).

The lymphocyte-predominant cells are considered to be the tumor cells in NLPHL, and they are not found in other HL subtypes. These cells are sometimes also called ‘popcorn cells’ because of the large folded or multilobed nuclear morphology (Mason *et al.*, 1994). Lymphocyte predominant cells constitute only about 1% of the cells in the tumor, the rest being infiltrating normal polyclonal inflammatory cells, such as follicular dendritic cells and macrophages, organized in a germinal-center like formation (Swerdlow *et al.*, 2008; Kuppers, 2009).

Histopathologically NLPHL may resemble other lymphomas, especially T-cell rich B-cell lymphoma and lymphocyte-rich classical HL, and thus immunohistochemistry is often used to facilitate the diagnosis (Swerdlow *et al.*, 2008; Smith, 2010). Commonly used immunohistochemical stainings include CD20, CD45 and CD79, CD30, CD15, and CD3 (Smith, 2010). Lymphocyte predominant cells are generally CD20, CD45, and CD79-alpha positive, but in contrast to the tumor cells in classical HL, they are commonly CD30 and CD15 negative. CD3 stains the normal T-cells that usually surround the lymphocyte predominant cells (Swerdlow *et al.*, 2008).

The etiology of NLPHL is poorly understood. Due to the rarity of NLPHL, the epidemiological studies usually discuss HL as one entity. In contrast to certain other subtypes of HL, previous infection with EBV is considered not to be relevant in NLPHL (Anagnostopoulos *et al.*, 2000). An increased risk of NLPHL has been reported in the first-degree relatives of NLPHL patients, suggesting shared genetic or environmental factors (Saarinen *et al.*, 2013). A few patients with autoimmune related lymphoproliferative syndrome, caused by germline mutation in *FAS*, have also been reported to have developed NLPHL (Sneller *et al.*, 1997; van den Berg *et al.*, 2002).

7.3 Meningioma

Meningiomas are the most common primary tumors of the central nervous system with incidence of approximately 4 and 9 per 100 000 person-years in the United States, in males and in females, respectively (CBTRUS Statistical Report, 2012). Meningiomas occur most commonly in individuals over 50 years of age, with a peak in the sixth and seventh decades of life (Louis *et al.*, 2007). In Finland, the World Standard Population age-adjusted incidence of meningiomas, is 1.6 in males and 5.5 in females, per 100 000 individuals (Larjavaara *et al.*, 2008). In Finland, as well as in many other industrialized countries however, the incidence of meningiomas has been rising during the past few decades (Klaeboe *et al.*, 2005; Larjavaara *et al.*, 2008).

Meningiomas originate from the meninges, which are the membranes surrounding the central nervous system, and arachnoidal cap cells are thought to be the origin of the tumorous cells. Great majority of meningiomas (90%) are found intracranially; the rest are located around the spinal cord. Multiple lesions are observed in 4-10% of the patients (Nahser *et al.*, 1981; Locatelli *et al.*, 1987; Antinheimo *et al.*, 2000; Huang *et al.*, 2005).

The biological spectrum of meningiomas is wide, and meningiomas are divided into a number of different histological subtypes. The most common are meningothelial, fibrous, and transitional subtypes, and these are usually benign and slowly growing (grade I) tumors. Atypical (grade II) and anaplastic (grade III) features are seen in less than 10% of all meningiomas (Louis *et al.*, 2007). Although meningiomas seldom metastasize, they can cause serious neurological symptoms by pressing adjacent structures. However, they can also be asymptomatic and go entirely unnoticed and up to 3% of meningiomas are found only at autopsy (Krampla *et al.*, 2004; Vernooij *et al.*, 2007).

The main treatment option for meningiomas is surgical removal, but the recurrence rate is high and up to 20% of the grade I tumors recur after successful resection (van Alkemade *et al.*, 2012). The grade II and III meningiomas are generally more aggressive, and they are associated with shorter survival times and higher recurrence rates of 29-52% and 50-94%, respectively (CBTRUS, 2012).

Risk factors for meningioma include ionizing radiation, and certain hereditary tumor suppressor syndromes are associated with increased risk of meningioma, including Neurofibromatosis type 2 (MIM #101000), Cowden syndrome (MIM #601728), and Werner syndrome (MIM #277700), caused by mutations in *NF2*, *PTEN*, and *RECQL2*, respectively (Flint-Richter and Sadetzki, 2007; Gorlin, 1987; Lauper *et al.*, 2013; Lloyd and Evans, 2013; Robinson and Cohen, 2000). In addition to these, a couple of families with meningiomas and a germline *SMARCB1* mutation, usually predisposing to familial schwannomatosis and rhabdoid tumors (MIM #162091 and 609322), have been described (Bacci *et al.*, 2010; Christiaans *et al.*, 2011; Melean *et al.*, 2012; van den Munckhof *et al.*, 2012). More recently, inherited loss-of-function mutations in *SMARCE1* were identified to predispose to familial clear cell meningiomas (Smith *et al.*, 2013; (Smith *et al.*, 2014)).

AIMS OF THE STUDY

This study was conducted to identify novel familial cases with unknown tumor predisposing conditions and to study genetic predisposition in them.

The more specific aims were:

1. To assess familial aggregation of different tumor types in Finland and to identify novel tumor predisposition families through systematic search in the Finnish Cancer Registry database
2. To identify susceptibility genes and elucidate their functions in families with Kaposi sarcoma, nodular lymphocyte predominant Hodgkin lymphoma and multiple intracranial meningioma

MATERIALS AND METHODS

Detailed information of the study subjects, samples and methods are described in the original articles (I-IV). Manufacturer's instructions and protocols have been followed if not otherwise indicated.

1. Ethical considerations

The studies were approved by the Ministry of Social Affairs and Health in Finland and by the local ethics review committee of the Hospital District of Helsinki and Uusimaa (HUS). Informed consent was obtained from the patients from whom fresh sample materials were collected. The use of archival formaline-fixed paraffine-embedded (FFPE) samples was authorized by the National Supervisory Authority for Welfare and Health (Valvira).

2. Study subjects and samples

2.1 Patients in the Finnish Cancer Registry (I)

In study I, all neoplasms registered in the Finnish Cancer Registry (FCR) in 1953-2011 were initially considered. This included 962 355 cancerous and 212 685 non-cancerous lesions, classified based on the International Classification of Diseases for Oncology 3rd edition (ICD-O-3) (World Health Organization, 2000). The non-cancerous lesions registered in FCR include basal cell carcinoma of skin, polycythemia vera, carcinoma in situ of breast, carcinoma in situ of bladder, non-invasive neoplasm of cervix uteri, borderline tumor of ovary, and papillary urothelial neoplasm of low malignant potential.

For the purpose of the clustering, the tumors were classified into 55 topography and 94 morphology groups, respecting the pathological similarities of different morphologies and topographies in the ICD-O-3 (Tables S1 and S2 in the online supplementary, study I). All possible topography-morphology combinations were considered, and combinations with at least two patients in FCR were used to define a tumor type in the clustering. Altogether 878 493 patients were linked to the National Population Registry (NPR) using identity numbers to obtain information on their family names at birth and birth municipalities.

2.2 Kaposi sarcoma patients (II)

Figure 1 summarizes the samples available from the Finnish KS family (study II). Additional samples used in study II included: nine Finnish familial KS cases, that belonged seven KS families (with two affected individuals) identified in study I, 56 sporadic Finnish KS patients, DNAs of previously reported Israeli cKS family from Israel (Guttman-Yassky *et al.*, 2004), and 26 familial cKS cases from Italy.

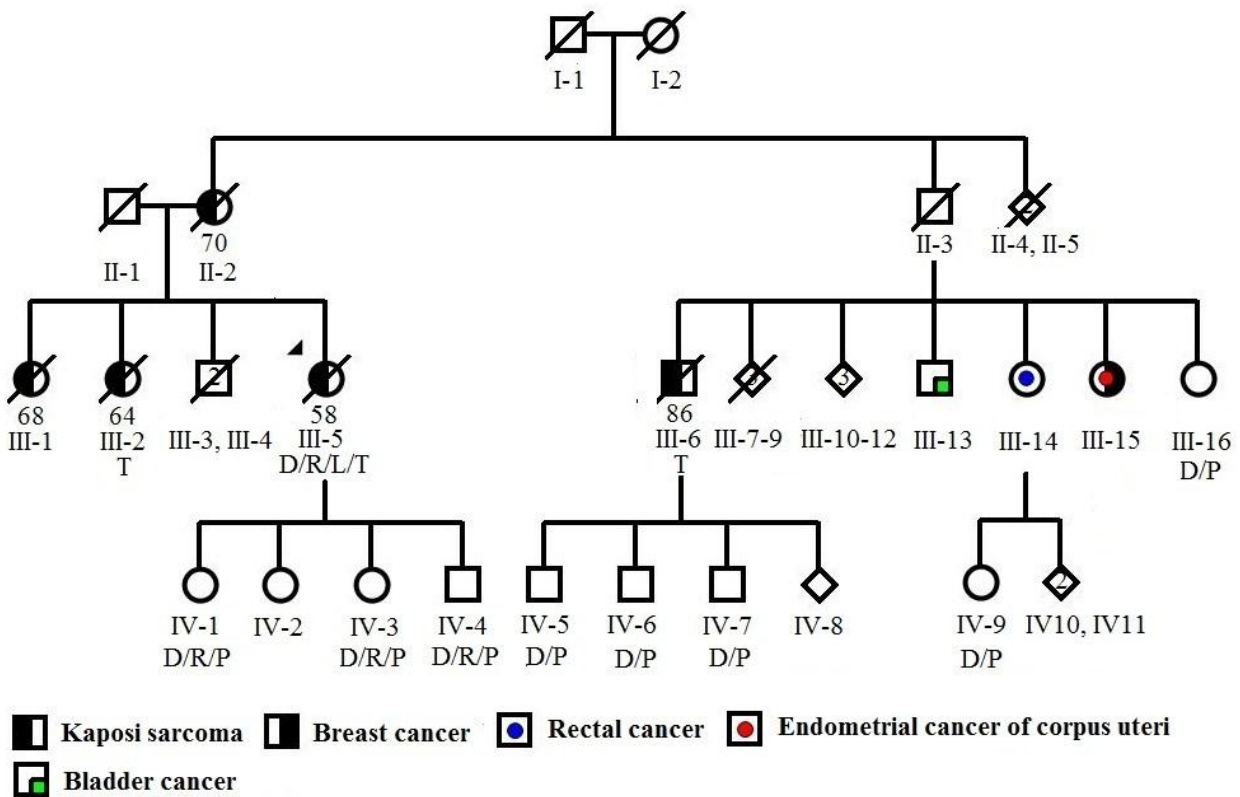


Figure 1. The Finnish Kaposi sarcoma (KS) family with five affected individuals in two generations. Round symbols denote females and square symbols males. Proband is indicated with an arrow and line through indicates a deceased individual. Numbers under the symbols denote the age at the KS diagnoses and letters the sample materials available (D: genomic DNA, T: DNA from formalin fixed paraffin embedded tumor tissue, R: total RNA from peripheral blood, P: peripheral blood mononuclear cells, L: EBV-transformed lymphoblast cell line). The pedigree has been slightly modified for confidentiality.

2.3 Hodgkin lymphoma patients (III)

Peripheral blood samples from 11 family members of the previously described Finnish NLP HL family (Saarinen *et al.*, 2011) were collected and used for extracting genomic DNA and constructing EBV-transformed lymphoblastoid cell lines (Figure 2). Additional peripheral blood samples were also collected from ten family members (Figure 2) for RNA extraction and transcriptome analysis. FFPE tumor samples from the four affected family members were also collected and these were used for confirming the lymphoma diagnoses as well as for EBV-testing (Figure 2).

Additional samples used in the *NPAT* screening included germline DNA samples from three previously published NLP HL families from UK (Indian; two affected siblings, Campbell *et al.*, 2004), France (Basque; father and two sons, Bauduer *et al.*, 2005) and Turkey (mother and child, Unal *et al.*, 2005), 73 FFPE tumor tissue DNA from Finnish HL patients, germline DNA samples

from 93 HL cases from UK, and NLPHL-derived DEV cell line (Atayar *et al.*, 2006). The 73 Finnish HL cases included 27 familial cases (with at least 2 HL cases, 2 NLPHL cases, or one HL and one NLPHL case in the family), 38 early onset NLPHL cases (less than 30 years at the time of diagnosis), and 8 NLPHL cases from the same geographical region as the NLPHL-Family 1. The HL cases from UK included 26 patients with at least one relative affected by a lymphoproliferative disorder and 67 sporadic HL cases.

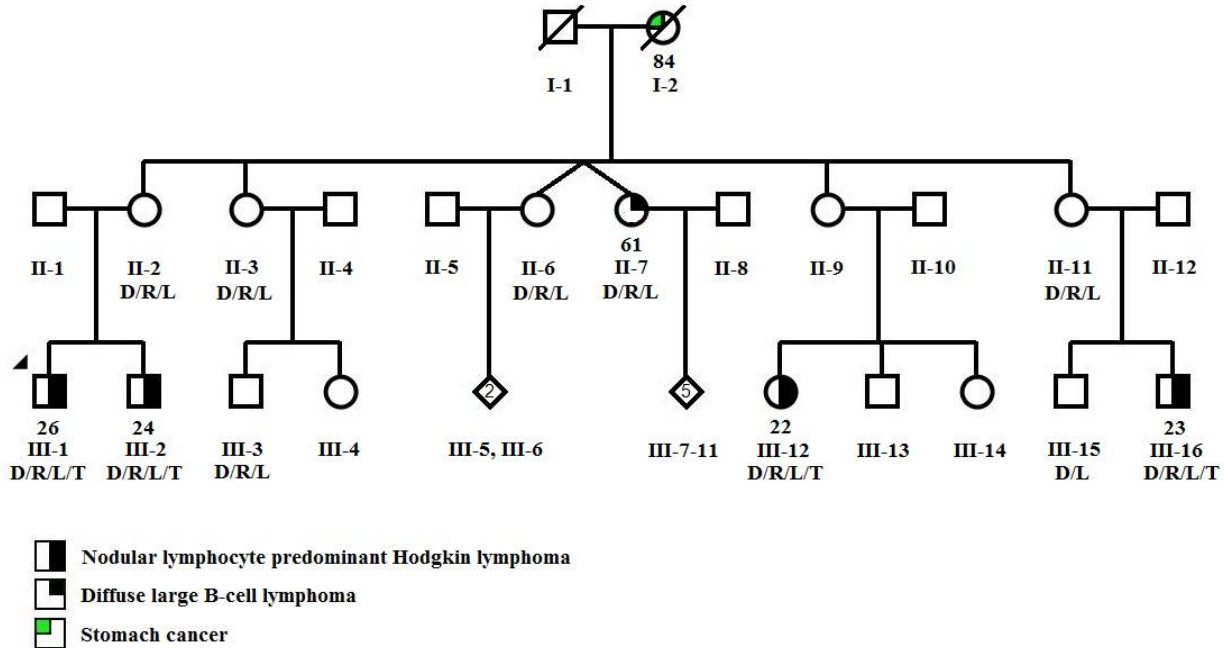


Figure 2. The Finnish nodular lymphocyte predominant Hodgkin lymphoma (NLPHL) family with four affected cousins. Round symbols denote females and square symbols males. The ages at cancer diagnoses are marked underneath the symbols. The index case is marked with an arrow and deceased individuals are marked with a line through. The letters under the symbols denote the sample materials available (D: germline DNA, T: DNA from formalin fixed paraffin embedded tumor tissue, R: total RNA from peripheral blood, L: EBV-transformed lymphoblast cell line). The pedigree has been slightly modified for confidentiality.

2.4 Meningioma patients (IV)

The family of five meningioma patients was initially identified by a clinician Pia Alhopuro. The samples and sample types available from the Finnish meningioma family members are marked in Figure 3. Additional samples used in *SUFU* screening consisted of germline DNA samples of 77 meningioma patients that had participated in the Interphone study (INTERPHONE Study Group, 2010), 35 *NF2* mutation negative multiple meningioma patients from UK, 5 familial meningioma patients (≥ 2 first-degree relatives with meningioma) from UK, and 4 Finnish *NF2* mutation negative patients with multiple meningiomas. In addition, 41 Finnish FFPE meningioma samples, of which 6 were from familial Finnish meningioma patients (≥ 2 first-degree relatives with meningioma), 20 from putative familial meningioma cases (identified in the clustering effort in

study I), 6 from meningioma patients from the same geographical region with the Finnish meningioma family, and 9 from patients with early onset meningioma (≤ 35 years), were also used for *SUFU* screening.

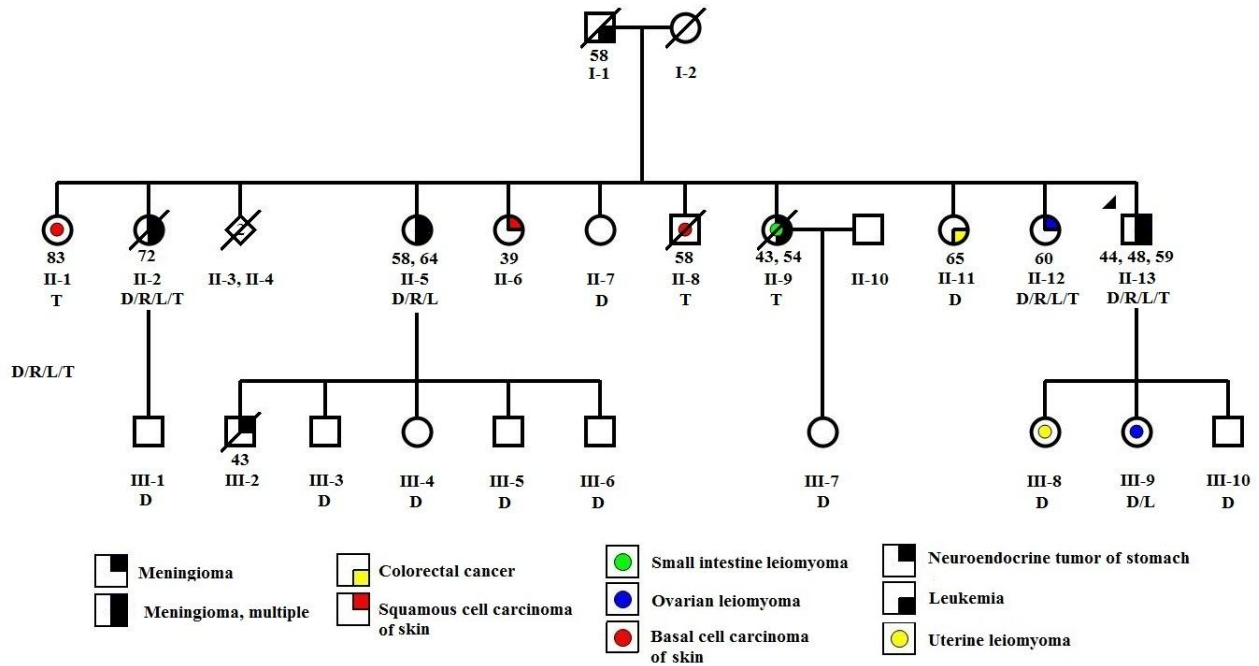


Figure 3. Pedigree of the Finnish meningioma family with five affected individuals. Round symbols denote females and square symbols males. The proband is indicated with an arrow and line through indicates a deceased individual. Numbers under the symbols denote the age at diagnoses and letters the sample materials available (D: genomic DNA, T: DNA from formalin fixed paraffin embedded tumor tissue: R, total RNA from peripheral blood, L: EBV-transformed lymphoblast cell line). The pedigree has been slightly modified for confidentiality.

2.5 Control samples (II, III, and IV)

Varying numbers of DNA samples extracted from peripheral blood of eligible Finnish blood donors were used as population specific controls in the studies II, III, and IV. The blood samples were acquired from the Finnish Red Cross Blood Transfusion Service. In study III, additional samples from healthy control individuals (n=177) from UK, participating in the National Study of Colorectal Cancer Genetics (Penegar *et al.*, 2007), were also used.

To filter the common polymorphisms from the high-throughput sequencing data, we used 149 Finnish genomes (in study II) and 13 and 78 Finnish exomes (in studies III and IV, respectively) sequenced for purposes of other projects in our laboratory. In study II, Finnish cohort of 1000 genomes (n=93, <http://browser.1000genomes.org/>), was used for additional controlling. In studies III and IV, additional controlling was performed using variants reposited in the dbSNP (<http://www.ncbi.nlm.nih.gov/SNP/>).

2.6 DNA and RNA extractions (II, III and IV) and cDNA synthesis (II and III)

The genomic DNA samples were extracted from peripheral blood and lymphoblastoid cell lines using the non-enzymatic DNA extraction method (Lahiri and Nurnberger, 1991) or DNeasy

Blood & Tissue kit (Qiagen). The genomic DNA from FFPE tissue sections was extracted either by standard phenol-chloroform protocol with proteinase K treatment or by Nucleo Spin FFPE DNA kit (Macherey-Nagel). The total RNA was extracted from peripheral blood using PAXgene Blood RNA kit (Qiagen) or from lymphoblastoid cell lines with RNeasy kit (Qiagen) or PureLink RNA Mini kit (Life Technologies).

Random primers (Promega), Moloney Murine Leukemia Virus (M-MLV) reverse transcriptase buffer (Promega), Deoxynucleotides (dNTPs; Bioline or Finnzymes), RNAase inhibitor (Promega), M-MLV enzyme (Promega), and 200-1000ng of total RNA was used in the cDNA synthesis. The random primers were annealed at 65°C (5min) and the reverse transcription was performed by incubating the reactions at 42°C for 50min followed by incubation at 95 °C for 10min.

3. Clustering of the patients in the Finnish Cancer Registry (I)

3.1 Clustering procedure (I)

In study I, two clustering efforts were performed. One was based on tumor type, municipality of birth, and family name at birth (MN-clusters) of the patients and the other on tumor type and family name at birth only (N-clusters). The observed (O) number of the patients for the different tumor types in FCR was calculated in a stratum specified by municipality, family name at birth, sex, and year of birth. Proportion of persons, in each characterized strata reflecting the entire Finnish population, was obtained from NPR. When these proportions were multiplied with the total sex and birth-year defined number of cancer cases in FCR, the expected (E) number of patients was defined. The stratum-specific O and E values were added up over the gender and birth year categories and for each cluster O/E ratios were calculated with 95% confidence intervals, assuming a Poisson distribution of observed numbers.

3.2 Cluster score calculation (I)

To calculate the cluster score, we selected MN-clusters with lower limit confidence interval (CI-low) ≥ 10 . The number of individuals in these was divided by the total number of patients with the same tumor type per 100 000 persons in Finland. To estimate if the obtained cluster score was different than the expected ratio two-sided Poisson test was calculated with 95% confidence interval. The p-values were adjusted for multiple testing with false discovery rate (Benjamini and Hochberg, 1995).

4. Genetic analyses

4.1 Genotyping (II, III and IV)

Genome-wide SNP genotyping was performed on Affymetrix 50K Xba SNP array (Affymetrix) (study III), Human610-Quad DNA analysis BeadChips (Illumina) (studies II and IV), or HumanOmniExpress-FFPE Bead Chip (Illumina) (study II). The sample preparation, hybridization, and scanning was performed at the Institute for Molecular Medicine Finland (FIMM) Genome and Technology Center. BeadStudio or GenomeStudio (Illumina) was used to call the genotypes and check the genotype qualities.

In study III, finemapping of the the linked regions was performed by genotyping microsatellite markers that were identified using Tandem repeat finder (Benson, 1999) or the Ensembl database (www.ensembl.org). The markers are listed in the online supplement (Table S1) of the original article (study III). The PCR products were run on ABI3730 capillary electrophoresis device (Applied Biosystems) at FIMM Technology Center, and the results were scored using GeneMarker software (SoftGenetics).

4.2 Transcriptome analysis (III and IV)

For the transcriptome analyses, the total RNAs were hybridized on a GeneChip Human Genome U133 Plus 2 (study III) or GeneChip Human Exon 1.0 ST Array (study IV) (Affymetrix). The sample preparations, hybridizations, and chip scanings were done at Biomedicum Functional Genomics Unit (FuGU) (Helsinki, Finland). Prior to normalization, the probes were annotated using Brainarray custom CDF file (ENTREZG, version 11.0.1) (Dai *et al.*, 2005). The data was normalized using robust multichip average normalization in R, using affy package. Fold change and two-tailed *t*-test with the equal variance assumption were calculated to evaluate *NPAT* and *NF2* expression differences in studies III and IV, respectively.

4.3 DNA copy number analysis (IV)

Human genome Comparative Genomic Hybridization (CGH) Microarray 105A (Agilent) was used to study the genome wide copy number variation (study IV). The genomic DNA from three affected female patients (II-2, II-5 and II-12; Figure 3) were pooled and hybridized on one array, and the genomic DNA from the affected male patient (II-13; Figure 3) was hybridized separately on another array. The sample preparation, hybridization, and scanning was done at FuGU. Data was analyzed and the CNVs called using circular binary segmentation algorithm to segment log R ratios into regions of estimated equal copy number (Olshen *et al.*, 2004).

4.4 Linkage analysis (II, III, and IV)

Multipoint linkage analyses were performed using Merlin (Abecasis *et al.*, 2002). In studies II and IV, Merlin's error detection algorithm was used to mark unlikely genotypes and pedwipe command to remove the problematic genotypes. In study IV, X chromosomal genotypes were analyzed separately with X chromosome specific version of Merlin (MINX).

In study II, genotypes were produced from the affected cousins (III-5 and III-6; Figure 1). SNPs shared between the two different platforms with minor allele frequency (MAF) ≥ 0.1 were selected for the linkage analysis. MAFs were determined using the CEPH population (Utahns originated from northern and western Europeans) in the HapMap phaseII data set (International HapMap Consortium *et al.*, 2007). A parametric linkage analysis with dominant inheritance model was performed, and consecutive markers with a positive LOD score and within 1cM distance from each other were considered.

In study III, both recessive and dominant parametric linkage analysis with 'affected only' strategy, with 100% penetrance, 0% likelihood of phenocopies, and in which healthy family members were coded as unknown for the disease status, was performed. Haplotypes were constructed using Merlin 'best' function.

In study IV, non-parametric linkage analysis was used. Individuals with multiple meningiomas (II-2, II-5, and II-13; Figure 3) were coded as affected, siblings without meningiomas as unaffected (II-7 and II-11; Figure 3), and the sibling with a single meningioma (II-12; Figure 3), and the three children of the affected family members (children of II-2, II-5, and II-7; Figure 3) as unknown. Allele frequencies were generated from the genotypes of the family and from 265 healthy Finnish control individuals provided by the Nordic Center of Excellence in Disease Genetics consortium.

4.5 Exome and genome sequencing (II, III, and IV)

The details of the exome capture kits, exome and genome sequence analysis tools, and analysis criteria are summarized in Table 3. In study II, the germline genome of III-5 (Figure 1) was sequenced from peripheral blood derived DNA using Illumina's paired-end sequencing chemistry. Two size-selected libraries (~1000bp and ~500bp) were sequenced using Illumina HighSeq2000 at FIMM Technology Center. The genomic library preparation and sequencing of the FFPE tumor sample of III-6 (Figure 1), was performed as an Illumina service (Illumina Cambridge Ltd, Little Chesterford, Essex, UK). Paired-end sequencing was used in all high-throughput sequencing applications and the sequencing was performed using Illumina sequencing chemistry and machines either at Karolinska Institutet (Sweden) or FIMM Technology Center.

In study II, the exomes and genomes were processed with an in-house pipeline (Saarinen *et al.*, 2013) with additional recalibration step. The adapter sequences were trimmed from the sequence reads using a script, and the read alignment was performed using the Burrows-Wheeler Aligner (BWA, Li and Durbin, 2010). Duplicate reads were removed with Markduplicates (<http://picard.sourceforge.net>), and the read processing and variant calling was performed following the Genome Analysis Toolkit (GATK, McKenna *et al.*, 2010) best practices (<http://www.broadinstitute.org/gatk/guide/best-practices>). Variant qualities and filtering were performed using an in-house high-throughput sequencing data visualization and analysis tool, "RikuRator". The variant annotations were based on Ensembl 71 (released in April 2013).

To identify structural genomic variants (SVs) in study II, BreakDancerMax (v1.2) (Chen *et al.*, 2009) was used. To call a SV, a minimum of 2 discordant read pairs (-r 2) and alternative mapping quality of 35 (-q 35) were required. The called breakpoints were flanked with 1000bp region up- and downstream for further filtering. SVs present in 22 Illumina (detected with BreakDancerMax) and 27 Complete Genomics (Complete Genomics Inc., Mountain View, CA, USA) in-house control genomes were filtered from the data. The breakpoints that colocalized with Segmental Dups (UCSC track table, Bailey *et al.*, 2001) or Hi Seq Depth regions (UCSC track table, Pickrell *et al.*, 2011) were also filtered out.

In studies III and IV, the exome sequencing data analysis was performed using commercial NextGENe software (Softgenetics) with NCBI36/hg18 and GRCh37/hg19 as the reference genome, respectively.

Table 3. The high-throughput sequencing methods and analysis tools used in studies II, III and IV

	Study II			Study III	Study IV
	Exomes		Genomes		
Samples	III-5 and III-2; Figure 1	7 Finnish familial KS cases	III-5 and III-6 ^{***} ; Figure 1	III-1; Figure 2	II-2, II-5 and II-13; Figure 3
Exome kit	Agilent SureSelect Human All exon Kit v1.0	Agilent SureSelect Human All exon Kit v1.0 or XT HumanAllExon	-	Agilent SureSelect Human All exon Kit v1.0	Agilent SureSelect Human All exon Kit v1.0
Sequencing machine	Illumina Genome Analyzer II and HighSeq 2000	Illumina HighSeq 2000	Illumina HighSeq 2000	Illumina Genome Analyzer II	Illumina Genome Analyzer II
Read length	58bp or 82bp	82bp	100bp	82bp	58bp , 77bp or 82bp
Analysis tools	In-house alignment and variant calling pipeline, visualization and analysis tool RikuRator (unpublished)	In-house alignment and variant calling pipeline, visualization and analysis tool RikuRator (unpublished)	In-house alignment and variant calling pipeline, visualization and analysis tool RikuRator (unpublished) or Illumina pipeline ^{***}	NextGENe v1.94 (Softgenetics)	NextGENe v2.1 (Softgenetics)
Reference genome	GRCh37/hg19	GRCh37/hg19	GRCh37/hg19	NCBI36/hg18	GRCh37/hg19
Minimum read coverage	5	5	5	5	4
Percentage of variant allele	25 %	25 %	25 %	20 %	20 %
Controls	149 genomes and 1000 Genomes Finnish genomes (MAF \geq 0.005)	149 genomes and 1000 Genomes Finnish genomes (MAF \geq 0.005)	149 genomes and 1000 Genomes Finnish genomes (MAF \geq 0.005)	13 exomes and dbSNP variants	78 exomes and dbSNP variants
Analyzed regions	linked regions, exons and exon- intron junctions [*]	exons and exon-intron junctions [*] of <i>STAT4</i> associated genes ^{**}	linked regions, exons and exon- intron junctions [*]	linked regions, exons and exon-intron junctions [*]	linked regions, exons and exon-intron junctions [*]

^{*} two bps to intron from the exon start and end sites; *STAT4* interacting, regulating or regulated genes identified through the use of QIAGEN's Ingenuity Pathway Analysis (Supplementary Table 2, study II, and IPA®, QIAGEN Redwood City, www.qiagen.com/ingenuity); ^{***} the FFPE tumor sample of III-6 sequenced and processed at Illumina Cambridge Ltd, Little Chesterford, Essex, UK.

4.6 PCR and Sanger sequencing (II, III, and IV)

Primers for PCR were designed using Primer3 (Koressaar and Remm, 2007; <http://bioinfo.ut.ee/primer3-0.4.0/primer3/>) or ExonPrimer programs (<http://ihg.gsf.de/ihg/ExonPrimer.html>). PCRs were performed in 25µl volume using AmpliTag Gold (Applied Biosystems), Phusion[®] High-Fidelity (Finnzymes), or Expand long template PCR system's (Clontech) DNA polymerases. The PCR products were run on agarose gel and purified with ExoSAP-IT (USB Corporation) or illustra[™] ExoSTAR 1 enzymes (VWR International). The purified products were sequenced at FIMM Technology Center. Sequences were analyzed manually with Chromas (<http://technelysium.com.au>) and using Mutation Surveyor software (Softgenetics).

4.7 Loss of heterozygosity analysis (IV)

Loss of heterozygosity (LOH) was studied by PCR, followed by Sanger sequencing, and scored manually by comparing the allele peak intensities of the germline and tumor tissue samples. For each variant site at least three parallel PCRs were performed and sequenced separately. LOH was scored if majority of the sequences (2/3) showed reduced wild type allele intensity.

5. Protein analyses

5.1 Serological testing of human herpes virus 8 (I and II)

In study I, serological testing was performed in age and sex matched serum samples of 200 Eastern and 200 Western Finns obtained from the National Public Health Institute's Health 2000 cohort. In study II, the serological testing was performed from the plasma of the additionally collected c.1337C>T *STAT4* variant carriers and their age and sex matched wild-type controls (Figure 1). The serological analyses were done at HUSLAB (Helsinki, Finland) with direct HHV8 IgG antibody Enzyme Linked ImmunoSorbent Assay.

5.2 In silico predictions and modelling (II, III and IV)

PolyPhen2 (Adzhubei *et al.*, 2010) and/or SIFT (Kumar *et al.*, 2009) were used to predict the effects of amino acid substitutions in studies II, III, and IV. In studies II and IV, the three dimensional protein models were produced with Crystallographic Object-Oriented Toolkit (COOT, Emsley and Cowtan, 2004), and the figures were made with PyMOL (DeLano, 2002). In study II, structural alignments of STAT1 and STAT3 (Protein Data Bank entries 1BG5 and 1BG1) were used to produce the model, since no crystal structure exist on STAT4. The p.Thr446Ile mutation was modelled using the to the structure of STAT3 (1BG1). In study IV, the crystal structure of the human N-terminal domain of SUFU was used (Protein Data Bank entry 1M1L).

5.3 Flow cytometry (II)

Ficoll-Paque solution (GE Healthcare) was used to separate peripheral blood mononuclear cells (PBMCs), and the CD4⁺ T cells were isolated using CD4⁺ T cell Isolation Kit II (Miltenyi Biotec). For assessment of activated STAT4, the CD4⁺ T cells were washed and suspended in X-VIVO-15 culture medium (Lonza) with 2 mM of additional L-glutamin. To induce IL-12R

expression, cells were plated on 96-well U-bottom plates (Nunc) for 20 h with anti-CD3/CD28 beads (Dynabeads® Human T-Expander CD3/CD28, Invitrogen) at bead to cell ratio of 1:1. Alternatively, cells were left untreated and then stimulated with IFN- α . To activate STAT4, cells were stimulated for 30 min at 37°C with IL-12p70 (R&D Systems) at a final concentration of 20 ng/ml or 100 ng/ml, or IFN- α (PBL Interferon Source) at a final concentration of 1000 IU/ml and fixed for 20 min on ice in 1.75% (w/v) of paraformaldehyde (PFA). Next, ice-cold methanol was added to achieve a final methanol concentration of 90%. The suspension was incubated for 10 min on ice after which the cells were washed with washing-blocking buffer with 10% fetal calf serum (FCS) and 0.02% (w/v) sodium azide in phosphate-buffered saline (PBS) and suspended in buffer with PE-conjugated STAT4 (pY693) monoclonal antibody (clone 38/p-Stat4; BD Biosciences), or an isotype control diluted in washing-blocking buffer, and the suspension was incubated for 30 min at room temperature. A pre-mixed cocktail of anti-CD4 FITC (SK3) and anti-CD45RO APC (UCHL1) antibodies (BD Biosciences) was added simultaneously to allow the identification of naive and memory T-helper cells. Cells were then washed twice with washing buffer consisting of 5% FCS and 0.02% sodium azide in PBS, suspended in 1% PFA in PBS, and stored overnight at 4°C prior to analysis with FACSCalibur flow cytometer (BD Biosciences). FlowJo software (Tree Star, Inc.) was used for the analysis of flow cytometry data.

For intracellular cytokine measurements in T cells, Ficoll-isolated PBMCs were cultured in X-VIVO-15 culture medium with 2 mM of L-glutamin on 96-well U-bottom plates overnight, and 10 ng/ml of PMA with 1 μ g/ml of ionomycin was added to stimulate intracellular cytokines. After 1 h of incubation, 10 μ g/ml of brefeldin A was added and the cells were incubated for another 4 h. Cells were then harvested and washed with washing buffer. Anti-CD3 FITC (SK7, BD Biosciences), anti-CD8 PerCP (SK1, BD Biosciences), and for some tubes also anti-CD45RO APC antibodies were incubated for 20 min at room temperature, and the cells were washed twice with washing buffer. The cells were fixed and permeabilized using Fix/Perm and Perm/Wash solutions (BD Biosciences). Anti-IFN- γ PE (25723.11, BD Biosciences) anti-IL-4 APC (8D4-8, BD Biosciences), anti-IL-17A Alexa Fluor® 647 (eBio64DEC17, San Diego, CA, USA), or isotype control antibodies were used for cytokine detection. After washing with Perm/Wash buffer, cells were suspended in 1% PFA in PBS and stored overnight at +4°C, prior to FACSCalibur flowcytometer analysis. Prism 6 software was used for statistical analyses (GraphPad Software Inc.), and Mann-Whitney U test was used to compare the two groups.

5.4 Cloning and site-directed mutagenesis (IV)

The c.367C>T substitution was generated to a human *SUFU* cDNA cloned in pCMV5_Myc vector (Clontech) using the QuikChange® Site-Directed mutagenesis kit (Agilent). The primer sequences used in the site-directed mutagenesis were: 5'GTTTTGGTGGTTTTGGCTTTGAGTTGACCTTTTGTCTGAAGAG3' and 5'CTCTTCAGACAAAAGGTCAACTCAAAGCCAAAACCACCAAAAC3'. The mutated clones were transfected to XL1-blue (Agilent), or TOP10 (Life Technologies) *E.coli* competent cells. The mutant and the wild-type clones were Sanger sequenced to ensure that the constructs

did not harbor additional mutations. Plasmid DNA was extracted using QuickLyse Miniprep kit (Qiagen) or NucleoBond® Xtra Mini kit (Machrey-Nagel).

5.5 Cell culturing and production of ShhN conditioned media (IV)

Human rhabdomyosarcoma cell line (CRL-2061), 293ShhN cell line, and *Sufu* (-/-) mouse embryonic fibroblasts (MEFs) were cultured in Dulbecco's modified eagle medium (DMEM) with 10% FCS and antibiotics. For generation of ShhN conditioned medium, 293ShhN cells were changed to low serum medium and the medium collected after 3 days; the medium was cleared by centrifugation and stored at +4°C.

5.6 Luciferase assay (IV)

Transfections were performed in 12-well plates with FuGENE HD (Roche) using 500 ng of pCMV5_Myc vector, (Clontech) with human wild type or c.367C>T mutant *SUFU*, 475 ng of the luciferase reporter plasmid "SASAKI COOPER" (Taipale *et al.*, 2000), and 25 ng of the *Renilla* luciferase phRL-TK reporter plasmid (Promega) to monitor the transfection efficiency. One day (CRL-2061) or two days (*Sufu* -/- MEFs) after the transfections, the cells were changed to low serum (0.5% FBS) medium with or without ShhN medium (1:5 v/v). Subsequently, the cells were incubated for 2–3 days after which the luciferase expression was analyzed using Dual-Reporter Luciferase system (Promega) with Digene DCR-1 luminometer (Diagnostics). The assays were conducted in duplicate and repeated at least three times. To calculate relative luciferase activity, the firefly luciferase counts were divided by the *Renilla* luciferase counts.

5.7 Immunoprecipitation and western blot (IV)

CRL-2061 cell lysates were centrifuged with a microcentrifuge at full speed at 4°C for 30 min, and the supernatant was removed and incubated with anti-Gli1 (H-300) agarose (Santa Cruz) for 2h at 4°C, followed by washing three times with lysis buffer without protease inhibitors prior to addition of a SDS sample buffer. Protein samples were resolved on SDS-PAGE gels and transferred to polyvinylidene difluoride membranes. The following antibodies were used in detection: goat anti-Myc HRP (ab1261; Abcam) and rabbit anti-Gli1 (ab92611; Abcam), in 1:5000 and 1:1000 dilutions, respectively. Goat anti-rabbit HRP (1:2000) was used as a secondary antibody. The proteins were quantified by Single Dimensional Electrophoretic Gel Analysis program from the ImageJ software (v1.46g).

5.8 Immunocytochemistry (IV)

Human fibronectin coated glass coverslips (BD) were placed on 6-well plates, and an appropriate number of *Sufu* (-/-) MEFs or CRL-2061 cells were seeded on top one day before transfection. The transfection was done with 1 µg of Myc-tagged human wild-type *SUFU* or c.367C>T mutant *SUFU* expression plasmids using FuGENE HD (Roche). After two days, the media was changed to a low serum medium with or without ShhN conditioned medium (1:5 v/v), and the cells were incubated for another two days. Specific anti-Myc (Santa Cruz) and Alexa 594-conjugated anti-mouse (Molecular Probes) antibodies were used for staining. Vectashield H-1200 (Vector) was used for mounting. Images were recorded with a Zeiss Axioplan epifluorescence microscope system and images were labeled using Adobe Photoshop software.

RESULTS

1. Cancer registry based familial clustering of cancers in Finland

To explore familial accumulation of different tumor types in Finland, we performed a systematic clustering of 878 593 patients registered in FCR. Patients were clustered based on tumor type, municipality of birth, and family name at birth (MN-clustering), or tumor type and family name only (N-clustering). As a result, 25 910 MN-clusters and 12 695 N-clusters, representing 183 different tumor types and with at least two cancer patients in each cluster, were formed. All clusters fulfilled the basic criterion that the number of patients in the cluster (O) was higher than the expected (E), based on the distribution of family names and/or birth municipalities in Finland and with the lower confidence interval (CI low) greater than or equal to one. The largest number of clusters was produced by patients with basal cell carcinoma of the skin, which is the most common tumor type in FCR. These clusters represented about 15% and 8% of all MN- and N-clusters, respectively.

To estimate the tumor types that showed greater familiarity than others, we calculated a cluster score, which reflects the fraction of patients belonging to the most significant MN-clusters ($CI\text{-}low \geq 10$) compared to the number of patients with the same tumor type in FCR, proportioned to 100 000 persons in Finland. Many of the tumor types that produced high cluster scores represented rare tumor types often related to certain hereditary cancer syndromes, such as hemangioblastoma, medullary thyroid carcinoma, hepatic neuroendocrine tumors, and nephroblastomas, but included also tumor types with less well characterized genetic etiology such as Kaposi sarcoma (KS), small intestine neuroendocrine tumors, myelosclerosis, and squamous cancer of the outer female genitals (Table 4). Of the more common tumor types, with over 7000 registered patients in FCR, the highest cluster scores were observed for papillary thyroid adenocarcinoma (cluster score 0.56), chronic lymphatic leukemia (cluster score 0.48), and squamous cell carcinoma of the lip (cluster score 0.35) (Table 1 of the original article, study I).

1.1 Frequent familial clustering of Kaposi sarcoma in Finland

To evaluate the efficiency of the method to identify true relatives, we chose to study Kaposi sarcoma (KS) with thorough genealogy. KS scored the third highest in our cluster score assessment (cluster score 1.91). The genealogy work revealed that 70% of the MN-clustered KS cases were true relatives (Figure 4A and 4B), and we identified altogether eight KS families. Seven of the identified families had two affected siblings, but one had five affected individuals in two generations (Figure 1). In this family, two of the affected family members were identified from MN-cluster, one was identified from an additional N-cluster, and one was connected through genealogy work.

We also studied the geographical distribution of the KS-derived MN- and N-clusters in Finland and noticed that the clustered KS cases accumulated mainly to the rural areas of western Finland (Figure 4A and 4B). Next, we studied the overall distribution of all KS cases in Finland, and that

Table 4. The tumor types showing the strongest familial accumulation

Topography	Morphology	Cluster score	p-value [*]	No. of patients in FCR	No. of clusters ^{**}	No. of patients in clusters ^{**}	Known genetic etiology
central nervous system	hemangioblastoma	4.98	2.77E-14	141	5	13	<i>VHL</i>
thyroid gland	medullary carcinoma	3.55	7.86E-21	335	9	22	<i>RET</i>
skin	Kaposi sarcoma	1.91	2.05E-13	537	9	19	
pancreas	neuroendocrine carcinoma	1.40	2.75E-06	386	5	10	<i>MEN1</i> , <i>RET</i>
kidney	nephroblastoma	1.26	5.98E-06	428	5	10	<i>WT1</i>
small intestine	neuroendocrine carcinoma	0.98	1.19E-09	1154	10	21	
hematopoietic and reticuloendothelial system	myelosclerosis	0.69	2.75E-09	2186	14	28	
vulva and vagina	squamous cell neoplasms and carcinoma	0.63	1.37E-10	3092	17	36	
mesothelium	mesothelioma	0.62	3.49E-06	1750	10	20	<i>BAP1</i>
breast	mucinous and mucinous cystic tumor	0.57	2.45E-07	2564	13	27	

^{*}Poisson distribution; ^{**}with lower confidence limit (CI-low) ≥ 10

also showed striking aggregation to western Finland, with another aggregate in northeastern Finland (Figure 4C). To test if this was due to uneven distribution of HHV8 infection in Finland, we studied seroprevalence of HHV8 from 200 age and gender matched non-KS affected individuals from eastern and western Finland. The overall HHV8 seroprevalence was 1% (n=4), with two of the seropositive cases originating from Eastern and two from Western Finland, indicating equal distribution of the HHV8 infection between Eastern and Western Finland. However, due to low number of positive cases the result remains suggestive.

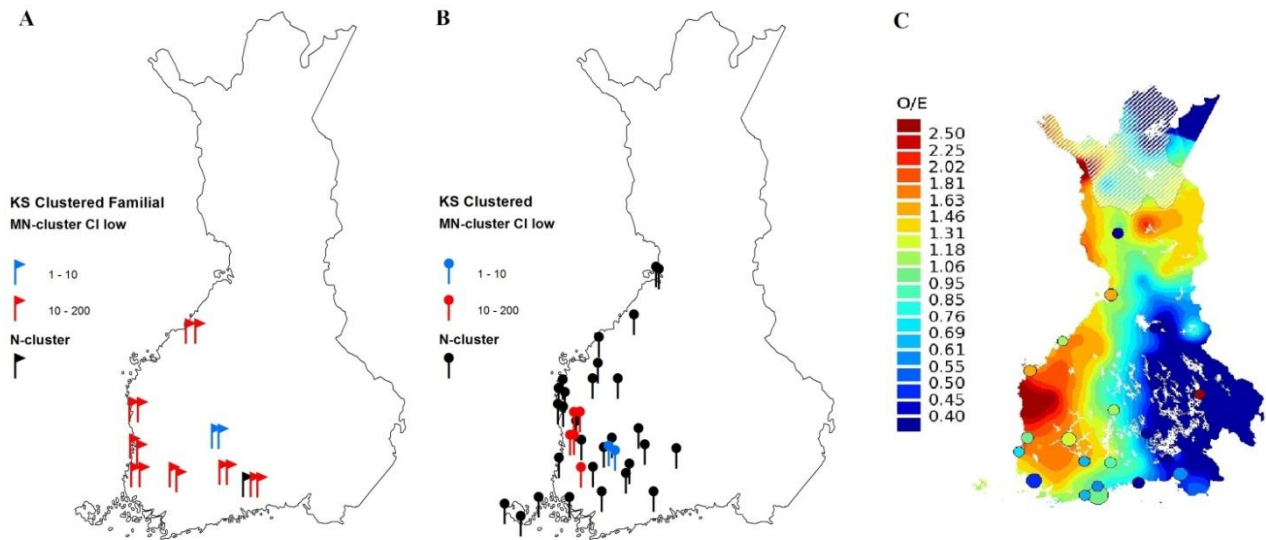


Figure 4. Geographical distribution of Kaposi sarcoma (KS) in Finland. **(A)** Clustered KS patients in Finland to whom we were able to confirm true kinship; each flag marks one KS patient; red and blue flags represent the lower confidence limit (CI-low) of MN-clusters, black denotes patients in N-cluster. **(B)** All other KS patients belonging to MN- and N-clusters in Finland. **(C)** The observed/expected (O/E) ratios of KS incidence according to municipality of birth; the largest cities are denoted with a circle with a diameter relative to expected number of cases and the color indicates the O/E; white stripes denote sparsely populated areas of Finland with <1 inhabitant per km^2 . Borders of Finland are based on those before year 1940 since many of the KS cases were born prior to that.

2. *STAT4* as a candidate susceptibility gene for Kaposi sarcoma

In study II, we studied the possible genetic predisposition in the Finnish cKS family with five affected individuals in two generations, originally identified as a part of study I. To exclude the chromosomal regions that were not shared by the affected family members, we performed genome-wide genotyping of the cousins (III-5 and III-6; Figure 1), followed by linkage analysis with dominant inheritance model. The chromosomal regions with negative LOD score and that were less than 1cM in length were considered not to be shared, and were excluded from further consideration. This resulted in 1350 cM, or 1.23 Gbp of shared chromosomal regions. The longest shared region was 89.7 cM in length and it was located at chromosome 2q14.2-2q35.

To identify the genomic variants that were located within the linked regions, we sequenced the genomes of the index and the affected cousin. After excluding intronic and non-synonymous coding variants, altogether 2646 shared protein coding variants remained. Subsequent filtering of common polymorphisms ($\text{MAF} > 0.005$ in 242 control genomes) reduced the list of shared candidate variants into 26, of which 14 segregated in the exome of the index case's sister (III-2; Figure 1 and Table 5). We also considered the structural variants (SVs) that the index case and the cousin shared, and we identified altogether 362 shared SVs, but none remained after filtering against the segmental duplications and the HiSeq Depth regions and the SVs present in the genomes of 49 in-house controls.

2.1 Prioritization of the candidate variants

Due to restricted sample materials, we were not able to study the segregation of the remaining 14 candidate variants in all five affected individuals in the family. Since immunity plays a central role in the development of KS we decided to concentrate on the candidate variants with known human immunity related functions. Among the 14 candidates, 3 resided in genes with known human immunity related functions including p.Val36Met in *HLA-C*, p.Pro212Ala in *HLA-DRB5* and p.Thr446Ile in *STAT4*. *In silico* prediction was used to study the putative effect of these variants. The variant in *HLA-DRB5* was predicted benign, whereas the variants in *HLA-C*, and *STAT4* were predicted to be damaging. Further search of these variants the public variant databases showed that the *HLA-C* variant (rs2308538) was not listed in the Exome Variant server or in the 1000 Genomes database, but it was found to be present with a relatively high population frequency in the dbSNP (MAF=0.2, dbSNP; build 134). The rs2308538 was also reported to map to seven alternative alleles of the human reference genome. The *STAT4* variant (rs141331848) on the other hand, was shown to be a very rare variant with global MAF of 0.0005 in the dbSNP as well as with MAFs of 0.001 and 0.0025 in the 1000 Genomes and Exome Variant server, respectively. We studied its presence in 180 population-matched blood donor controls originating from the same geographical region as the KS family and detected no variant carriers.

The c.1337C>T substitution encodes a missense change altering the amino acid 446 of *STAT4* from threonine to isoleucine (p.Thr446Ile). Threonine 446 is a highly conserved amino acid in evolution and it is located within the DNA binding domain of *STAT4* (see Figure 2B and C in the original manuscript, study II). The c.1337C>T substitution is also located two nucleotides from the 3' exon-intron boundary of the exon 16 in the *STAT4* (CCDS2310) and thus we hypothesized that it may affects the mRNA splicing of *STAT4*. We studied the possible splicing defect of the *STAT4* by cDNA sequencing, but the splicing was shown to be not affected.

Table 5. The segregating variants in the Finnish KS family

Gene	Genomic position	Nucleotide Change	Amino acid change	Genotype	rs-code	In healthy controls (n=242)
<i>RANBP2</i>	2:109374917	C->T	His839Tyr	Het	-	0
<i>CYP4F31P</i>	2:132044787	G->A	Val36Met	Het	rs112500726	1
<i>STAT4</i>	2:191904022	G->A	Thr446Ile	Het	rs141331848	0
<i>C3orf52</i>	3:111835543	G->A	Glu151Lys	Het	rs201307291	2
<i>MUC4</i>	3:195512212	G->T	Pro2080His	Het	rs75588776	0
<i>NUDT6</i>	4:123814164	G->A	Thr257Ile	Het	rs45606640	1
<i>SLC7A11</i>	4:139106315	A->G	Ile292Thr	Het	rs144628608	0
<i>CARD6</i>	5:40843521	G->A	Gly184Glu	Hom and Het	rs186662969	2
<i>HLA-C</i>	6:31239613	C->T	Val36Met	Het	rs2308538	1
<i>HLA-DRB5</i>	6:32487165	G->C	Pro212Ala	Het	rs142204283	1
<i>MS4A10</i>	11:60565924	T->C	Val220Ala	Het	rs139578350	2
<i>MTMR10</i>	15:31233884	T->C	Tyr708Cys	Het	rs200560796	0
<i>SPTBN5</i>	15:42170634	G->A	Arg1126Cys	Het	rs200015124	0
<i>ADAMTS1</i>	21:28210069	C->A	Gln911His	Het	rs200480524	0

Genomic positions are described based on the reference build GRCh37/hg19; rs-codes are from dbSNP Build 139. Het denotes heterozygous and Hom homozygous.

2.2 STAT4 activation is not affected by p.Thr446Ile mutation

To see if the p.Thr446Ile mutation affected STAT4 function, we induced naive and memory CD4⁺ T cells of the mutant and wild-type STAT4 carriers using IL-12 or IFN- α . We analyzed the activated cell populations from four STAT4 p.Thr446Ile carriers and eight age- and sex-matched wild-type controls by flow cytometry, using antibody against phosphorylated STAT4 (pY693). Comparable amounts of phosphorylated STAT4 were detected in the mutants and in the healthy controls, indicating that the heterozygous p.Thr446Ile mutation does not affect the phosphorylation of STAT4 (Supplementary data, Figures 1B and C of the manuscript, study II).

2.3 IFN- γ production is decreased in the p.Thr446Ile variant carriers

Next, we tested, if the IFN- γ -responses of the mutant carriers were affected. IFN- γ staining intensity, following PMA and ionomycin stimulation, was significantly decreased in the naive CD3⁺CD8⁻ T helper (Th) cells of the p.Thr446Ile variant carriers (Figure 3B of the manuscript, study II). Similar trend, although not statistically significant, was also present in the memory CD3⁺CD8⁻Th cells of the p.Thr446Ile variant carriers (Figure 3B of the manuscript, study II). No differences were found in the percentages of IFN- γ -positive cells, however (see Figure 3A of the manuscript, study II). The IFN- γ responses were also measured in the CD8⁺ cytotoxic T cells of the p.Thr446Ile variant carriers and wild-type subjects, but the responses remained at comparable levels (Figure 3B of the manuscript, study II).

2.4 IL-4 and IL-17 responses in the p.Thr446Ile carriers

We also assessed the IL-4 (a marker for Th 2 cells) and IL-17 (a marker for Th 17 cells) expression in the PMA and ionomycin stimulated CD4⁺ T cell populations and CD8⁺ T cells of the variant carriers and the wild-type individuals by intracellular flow cytometry. Similar amounts of IL-4 and IL-17 were observed in the naive and memory CD4⁺ T cell populations in both groups. In the CD8⁺ T cells, the expression of IL-4 and IL-17 remained at negative or very low level, with no detectable differences between the study groups.

2.5 Screening of *STAT4* and associated genes

To genetically validate *STAT4* as a KS predisposing gene, we analyzed the protein coding region of *STAT4* from the exomes of seven affected KS patients from six Finnish KS families, but did not identify any coding mutations. We also screened, by Sanger sequencing, the entire coding region of *STAT4* from 27 familial cKS patients, from Italy (n=26, belonging to 13 families) and Israel (n=1), but identified only common intronic variants. In addition to this, we sequenced the c.1337C>T variant site from 56 sporadic Finnish KS tumors samples. Two of the tumors harbored a rare synonymous coding variant (c.1338C>A, p.Thr446Thr; rs144421302) variant. Although the variant encodes a synonymous change it is also located at a putative splice site of the exon 16 of *STAT4*. The variant listed in public databases, and it is present in the Exome Variant Server and NCBI dbSNP with MAFs 0.0025 and 0.002, respectively.

Since we did not identify additional *STAT4* protein code altering variants, we hypothesized that KS susceptibility variants could also be identified from *STAT4* related genes. We generated a list of *STAT4* interacting, regulating or regulated genes (n=47, Supplementary Table 2 of the original manuscript, study II) through the use of QIAGEN's Ingenuity Pathway Analysis (IPA[®], QIAGEN Redwood City, www.qiagen.com/ingenuity), and studied the exomes of the seven Finnish familial KS cases for mutations in these genes. Rare coding variants (MAF<0.01) were identified in *TYK2* and *CCR5* (Table 6). We validated these by Sanger sequencing and their segregation was studied in the affected family members in respective families, and only the p.His289Arg variant in *CCR5* was identified to segregate with the KS in the family of two affected siblings. The variant is *in silico* predicted to be damaging, and it is located within a putative transmembrane region of *CCR5*.

Table 6. Variants in *STAT4* related genes

Gene	Genomic position	Nucleotide change	Amino acid change	Genotype	rs-code	In healthy controls (n=242)
<i>CCR5</i>	3:46415259	A->G	His289Arg	Het	-	0
<i>TYK2</i>	19:10472967	G->A	Arg548Cys	Het	-	0
<i>TYK2</i>	19:10478867	C->T	Arg110Gln	Het	rs56090991	0

Genomic positions are described based on the reference build GRCh37/hg19; rs-codes are from dbSNP Build 139. Het denotes heterozygous.

3. NPAT may predispose to nodular lymphocyte predominant Hodgkin lymphoma

3.1 Analysis of large chromosomal alterations and EBV-infection

In study III, a Finnish family with four NLPHL affected cousins (Saarinen *et al.*, 2011) was studied for a genetic predisposition. The possibility of a larger chromosomal alteration as the cause of NLPHL predisposition was excluded by standard G-banding from the index case (III-1; NLPHL-family, Figure 2), whose karyotype was shown to be normal. The tumors of the affected individuals were also tested for presence of EBV, by EBV-encoded RNA in situ hybridization, but no sign of EBV was detected.

3.2 Identification of *NPAT* as a candidate susceptibility gene

To identify putative predisposition genes, we performed genome-wide linkage analysis followed by exome sequencing. Genome-wide SNP genotypes were produced from all family members of the NLPHL family from whom genomic DNA was available. Next, we performed a parametric linkage analysis with 'affected only' approach. Thirteen genomic regions, spanning 112 Megabases (Mb) of the genome, segregated with NLPHL in the family (Table 2 of the original article, study III). The regions were fine-mapped with microsatellite markers, but none of the regions could be excluded based on this.

To identify the protein code altering variants in genes that resided within the linked regions, the index cases's germline was exome sequenced (III-1; Figure 2). After combining the variants from the exome with the linked regions, we removed all synonymous, intronic, and bad quality variants as well as the variants present in 13 in-house control exomes. This resulted in 14

heterozygous candidate variants. These were further confirmed by Sanger sequencing, and half of the variants were shown to be false positive calls. Of the remaining seven variants, six segregated in all four affected family members (Table 7). These were studied in additional Finnish blood donor samples, and the only variant that was not present in the controls was a heterozygous frameshift mutation (c.2437-2438delAG) in the *nuclear protein, ataxia-telangiectasia locus* (*NPAT*) (Table 7). The mutation destroys the reading frame and putatively creates a premature stop codon after five encoded amino acids (p.Leu814Phefs*5).

We further studied the presence of the c.2437-2438delAG *NPAT* mutation in seven unaffected family members (II-2, II-3, II-6, II-7, II-11, III-3 and III.15; Figure 2), and five were shown to carry the deletion. Three of the carriers were mothers of the affected individuals and thus they were obligatory carriers of the deletion. The other two mutation carriers were monozygotic twin aunts of the affected individuals (II-6 and II-7; Figure 2).

Table 7. Exomic variants segregating with nodular lymphocyte predominant Hodgkin lymphoma in the Finnish family of four affected individuals

Gene	Genomic Position	Nucleotide Change	Amino acid Change	Genotype	rs-code	In healthy controls
<i>SLC22A14</i>	3:38349147	delTGTT	p.Leu238Serfs*	Het	rs138257211	1/3
<i>NPAT</i>	11:108043273	delCT	p.Leu814Phefs*	Het	-	0/239
<i>KRT85</i>	12:52758810	C>T	p.Asp189Asn	Het	rs112554450	1/3
<i>DGKA</i>	12:56334185	A>G	p.Thr296Ala	Het	rs141417048	2/110
<i>MIP</i>	12:56848079	C>T	p.Val107Ile	Het	rs74641138	2/3
<i>POF1B</i>	X:84569452	G>A	p.Arg315Cys	Hom	rs76123916	2/251

Genomic positions are described based on the reference build GRCh37/hg19; rs-codes are from dbSNP Build 139. Het denotes heterozygous and Hom homozygous.

3.3 Expression of *NPAT* in the c.2437-2438delAG carriers

NPAT expression was studied from 8 mutation carriers and 12 controls by transcriptome analysis. The expression was shown to be reduced in the mutant carriers (fold change 0.80, p=0.016). This was validated by *NPAT* cDNA sequencing from nine *NPAT* c.2437_2438delAG carriers and two *NPAT* wild-type family members. The mutant allele peaks in all nine mutation carriers showed lowered peak intensities of the mutant allele (Figure 5). *NPAT* cDNA was also screened from NLPHL-derived DEV cell line as well as from the cDNAs of two previously reported NLPHL families (Campbell *et al.*, 2004; Bauduer *et al.*, 2005), but no mutations were identified.

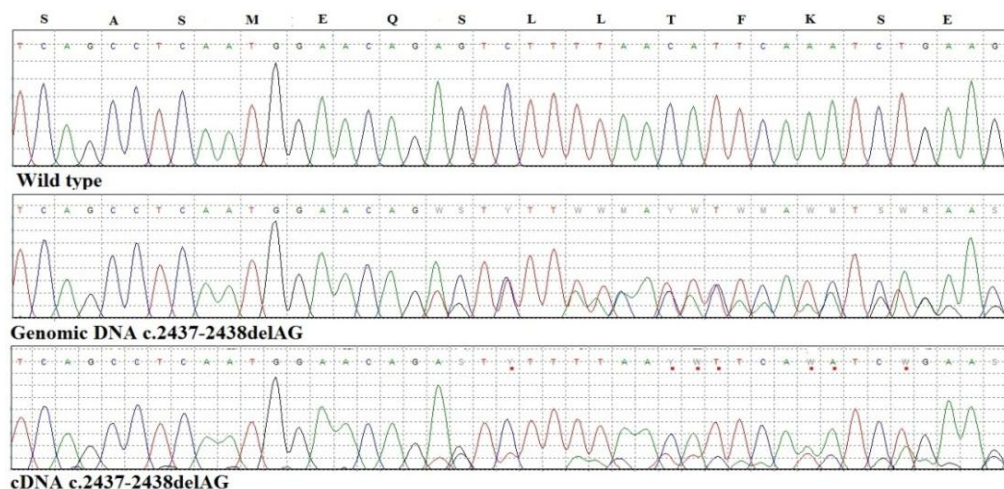


Figure 5. cDNA of *NPAT* wild-type individual and cDNA and genomic DNA of heterozygous *NPAT* c.2437-2438delAG mutation carrier. The reduced expression of the mutant allele can be seen in the cDNA of the mutation carrier.

3.4 In-frame deletion of p.Ser725 in Hodgkin lymphoma

To find additional patients with *NPAT* mutations, we screened the entire coding region and exon-intron boundaries of *NPAT* from 73 Finnish and 93 UK HL cases. A rare in-frame deletion (c.2171_2173delCTT), which deletes the serine at the amino acid position 725 (p.Ser725del) was identified from one Finnish NLP HL case and from six HL cases from UK. The Finnish NLP HL case was an early onset NLP HL patient, diagnosed at the age of 21. Four of the UK cases were also early onset NLP HL patients, diagnosed before the age of 30. The other two UK cases had a first degree relative with non-HL and HL, respectively. We studied the frequency of the deletion in the available healthy Finnish and UK control samples. Two Finnish and three UK controls were found to carry the deletion. On the basis of the observed p.Ser725del carriers, an odds ratio for p.Ser725del and HL was calculated. Odds ratio of 4.11 (with 95% CI, 1.27-13.35, $p=0.018$) was observed, suggesting an increased risk of HL in the mutant carriers.

4. Loss-of-function mutation in *SUFU* in meningioma predisposition

4.1 *NF2* is not associated with the Finnish meningioma family

In study III, we aimed to identify the possible genetic cause of the intracranial meningiomas in a Finnish family with five affected individuals. The index case (II-13; Figure 3) of the family was directed to genetic counseling after his third meningioma operation. His four sisters (II-2, II-5, II-9 and II-12; Figure 3) had also had meningiomas. Three out of the four sisters (II-2, II-5 and II-12; Figure 3) had had multiple meningiomas and one had had (II-9; Figure 3) a solitary meningioma.

Because multiple neoplasia syndrome Neurofibromatosis, type II, *NF2* (MIM #101000) is associated with increased risk of meningiomas, the index case was examined for germline mutations in *NF2* at the medical genetics clinic by Sanger sequencing and multiplex ligation-dependent probe amplification. However, no *NF2* mutations were detected. We also confirmed by transcriptome analysis that the patients and the control subjects had comparable expression of *NF2* (fold change=0.99) in the blood.

To explore the possibility that a larger copy number gain or loss would cause the meningioma susceptibility, we performed array comparative genomic hybridization from four affected family members. The studied individuals shared altogether 18 copy number alterations, but these were all shown to have multiple entries in the Database of Genomic Variants (<http://dgv.tcag.ca/dgv/app/home>) and were thus not considered to be relevant for the meningioma predisposition (Table 8).

Since about 60% of sporadic meningiomas harbor somatic *NF2* mutations (Louis *et al.*, 2007), we wanted to see if the tumors from the affected family members also harbored somatic *NF2* mutations. We Sanger sequenced the entire coding region and the exon-intron boundaries of *NF2* were from the seven available meningiomas belonging to four affected individuals, but identified no *NF2* mutations, indicating that the meningiomagenesis in the family was not related to *NF2*.

4.2 Identification of meningioma susceptibility genes

To identify novel meningioma susceptibility alleles, a combination of genome-wide SNP genotyping, linkage analysis and exome sequencing was used. Owing to the family structure (Figure 3), nonparametric linkage analysis was performed. The linkage analysis resulted in 174 chromosomal regions with positive LOD scores and with a total length of 1.3 Gbp of shared genomic regions (online supplement, Table S1 of the original article, study IV).

To study the genetic variants residing in the linked regions, we sequenced the germline exomes of three affected family members (II-2, II-5 and II-13; Figure 3). All synonymous coding variants, variants present in the 78 in-house control exomes, listed in dbSNP (v.132), and variants located outside the linked regions were excluded from further consideration. As a result, 20 protein code altering candidate variants remained (Table S2 of the original article, study IV). The segregation of these was further studied by Sanger sequencing in all affected individuals, and seven heterozygous missense mutations were found to segregate with the meningiomas in the family (Table 9).

Next, we studied the presence of these 7 variants in a set of 188 healthy Finnish blood donor samples and samples from healthy family members. Three of the variants were found to be present in the controls and two were found from unaffected siblings, resulting in two potential heterozygous candidates, the c.1990A>G (p.Ser664Pro) in *AASS* and the c.367C>T (p.Arg123Cys) in *SUFU* (Table 9). To assess the putative impact of the *AASS* p.Ser664Pro and the *SUFU* p.Arg123Cys, we performed *in silico* predictions with Polyphen2 (Adzhubei *et al.*, 2010) and SIFT (Kumar *et al.*, 2009). Both programs predicted the p.Arg123Cys *SUFU* mutation to be damaging (Figure 6A) and the *AASS* p.Ser664Pro to be benign. Multiple sequence alignment also showed that the arginine 123 of *SUFU* is highly conserved in evolution (Figure 6B). We then studied the presence of possible loss of heterozygosity (LOH) at the site of the *SUFU* and *AASS* variants. LOH was observed at the site of the *SUFU* mutation in all of the seven studied tumors but no LOH was observed at the site of the *AASS* mutation.

Table 8. The copy number alterations in the Finnish meningioma family

<i>Region</i>	<i>Patients</i>	<i>Chromosome</i>	<i>Start bp</i>	<i>End bp</i>	<i>No of markers</i>	<i>CNV</i>
1	II-13	1	72539172	72579340	14	Loss
	II-2, II-5, II-12	1	72539172	72579340	14	Loss
2	II-13	1	150836504	150851468	10	Loss
	II-2, II-5, II-12	1	150823099	150852881	12	Loss
3	II-13	1	157134825	157136371	2	Loss
	II-2, II-5, II-12	1	157134825	157136371	2	Loss
4	II-13	1	195062822	195150019	28	Loss
	II-2, II-5, II-12	1	195022883	195150019	31	Loss
5	II-13	2	33619289	33621140	2	Loss
	II-2, II-5, II-12	2	33619289	33621140	2	Loss
6	II-13	3	163997257	164027753	8	Gain
	II-2, II-5, II-12	3	163997257	164107980	28	Gain
7	II-13	5	12866067	12873081	2	Loss
	II-2, II-5, II-12	5	12866067	12873081	2	Loss
8	II-13	5	97074093	97119405	7	Loss
	II-2, II-5, II-12	5	97060234	97126947	9	Loss
9	II-13	6	32563281	32600991	9	Loss
	II-2, II-5, II-12	6	32558699	32600991	10	Loss
10	II-13	7	133436294	133448306	8	Loss
	II-2, II-5, II-12	7	133436294	133448306	8	Loss
11	II-13	7	141698663	141714197	10	Gain
	II-2, II-5, II-12	7	141700461	141714197	9	Gain
12	II-13	10	114103576	114106489	2	Loss
	II-2, II-5, II-12	10	114103576	114106489	2	Loss
13	II-13	13	33033934	33042586	3	Loss
	II-2, II-5, II-12	13	33033934	33042586	3	Loss
14	II-13	14	105602631	105632197	12	Loss
	II-2, II-5, II-12	14	105602631	105632197	12	Loss
15	II-13	16	57204888	57206801	2	Loss
	II-2, II-5, II-12	16	57204888	57206801	2	Loss
16	II-13	19	51316683	51319033	3	Loss
	II-2, II-5, II-12	19	51316683	51319033	3	Loss
17	II-13	20	1511661	1531191	9	Loss
	II-2, II-5, II-12	20	1511661	1531191	9	Loss
18	II-13	22	22677988	22725329	22	Gain
	II-2, II-5, II-12	22	22677988	22725329	22	Gain

Start bp and End bp are based on the human reference build NCBI36/hg18. Patient numbers refer those in Figure III.

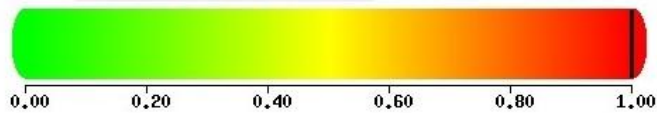
Table 9. Exomic variants segregating with meningiomas in the Finnish family of five affected individuals

Gene	Genomic position	Coding nucleotide	Amino acid change	In controls	In unaffected family members	rs-code
<i>CTRC</i>	1:15772126	c.697A>C	p.Glu225Ala	0/187	Yes	rs201486613
<i>ZNF80</i>	3:113955753	c.673C>T	p.Val57Met	0/182	Yes	rs146464395
<i>AASS</i>	7:121731783	c.2135A>G	p.Ser664Pro	0/174	No	rs202247760
<i>SUFU</i>	10:104309776	c. 367C>T	p.Arg123Cys	0/180	No	rs202247756
<i>IFIT2</i>	10:91067047	c.1509A>G	p.Lys445Arg	3/69	nd	rs138979563
<i>KL</i>	13:33628244	c.1168G>A	p.Arg387His	1/90	nd	rs200112661
<i>SERPINA12</i>	14:94953767	c.1914G>A	p.Ala373Val	1/89	nd	rs138209155

Genomic positions are based on the human reference genome build GRCh37/hg19; rs-codes are adopted from the dbSNP Build 139, nd: not determined.

A

This mutation is predicted to be **PROBABLY DAMAGING** with a score of **1.000** (sensitivity: **0.00**; specificity: **1.00**)



B

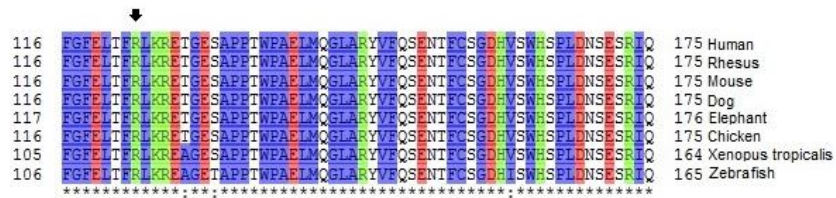


Figure 6. *In silico* assessment of p.Arg123Cys SUFU mutation. (A) Polyphen2 prediction of the p.Arg123Cys mutation. (B) Multiple sequence alignment of SUFU. The p.Arg123 is marked with an arrow. Blue, red and green color denotes hydrophobic, negative, and positive amino acids, respectively.

4.3 *In silico* modelling of the p.Arg123Cys SUFU mutation

To investigate effect of the p.Arg123Cys SUFU mutation further, we performed three-dimensional modelling of the mutant and wild-type SUFU using Crystallographic Object Oriented Toolkit (COOT, Emsley and Cowtan, 2004). This implicated that the arginine 123 is located in the center of the N-terminal subdomain of SUFU and that it forms multiple hydrogen bonds and ionic interactions with histidine 89, aspartic acid 182 and glutamines 184 and 199 (Figure 7A). These interactions are central for the loop structures of the N-terminal subdomain of SUFU. When the arginine 123 is substituted by cysteine, the interactions are lost and the loop structures are relaxed, possibly affecting the proper tertiary structure of SUFU (Figure 7B).

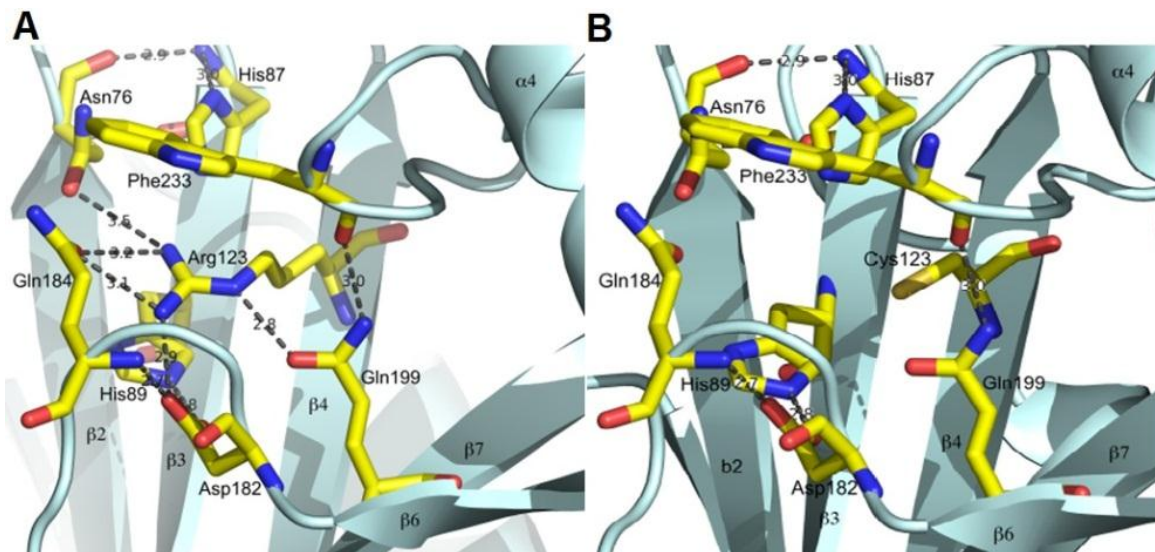


Figure 7. Computational modelling of the SUFU p.Arg123Cys mutation.

(A) Ribbon representation of wild-type SUFU. The p.Arg123 and the interacting residues are illustrated as balls and sticks. Dashes represent hydrogen bonds and ionic interactions. The secondary structure elements are labeled in accordance to Merchant *et al.*, 2004. (B) The p.Arg123Cys mutant SUFU. All hydrogen bonds and ionic interactions are lost possibly interfering with the formation of the proper tertiary structure of the protein and thus its function.

4.4 Mutated *SUFU* leads to dysregulated Hedgehog signaling

To elucidate the properties of the p.Arg123Cys SUFU mutation *in vitro*, we studied its ability to suppress GLI1 in human rhabdomyosarcoma cells (CRL-2061). These cells have constitutively activated Hedgehog signaling (Hh) pathway due to 30-fold amplification of *GLI1* (Roberts *et al.*, 1989). The cells were transfected with the mutant or wild-type *SUFU* expression construct with Hh-specific reporter plasmid and a control reporter plasmid. The luciferase activity was measured after 2-3 days of incubation. The assays were conducted with and without ShhN-conditioned medium in duplicate and were repeated at least three times. The relative luciferase activity was calculated by dividing the firefly luciferase counts by the Renilla luciferase counts. Compared to the wild-type SUFU, the mutant SUFU had significantly reduced ability to suppress Hh activity (Figure 8A). Comparable results were obtained also when the same experiments were performed in *Sufu* knock-out (*Sufu*^{-/-}) MEFs (Figure 8B).

SUFU is known to directly interact with GLI1 and regulate its shuttling from nucleus to cytoplasm (Kogerman *et al.*, 1999). To test if the reduced suppressive activity of the mutant SUFU was due to impaired interaction with GLI1, we tested their interaction in the mutant and wild-type SUFU transfected CRL-2061 cells with a pull-down immunoblot analysis using primary antibodies against Myc-tagged SUFU and GLI1. Compared to the wild-type SUFU, the protein level of the p.Arg123Cys SUFU was decreased and it was found to bind clearly less GLI1 (Figure 8C).

To further elucidate if the p.Arg123Cys mutation affected the subcellular localization of GLI1, we transfected CRL-2061 cells and *Sufu*^{-/-} MEFs with Myc-tagged wild-type and mutant SUFU expression plasmids. The cells were incubated two days with or without ShhN-conditioned

medium after which they were stained with antibodies against GLI1 and Myc-tagged SUFU. In the mutant SUFU transfected cells, GLI1 was localized more in the nuclei, indicating impaired re-localization of GLI1 to the cytoplasm (Figure 9).

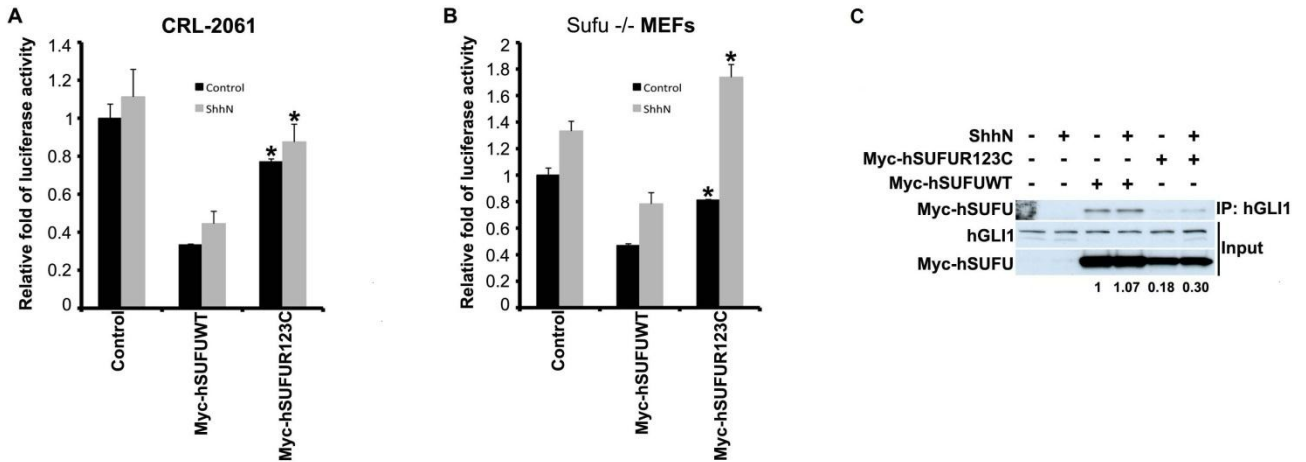


Figure 8. The p.Arg123Cys mutated SUFU has reduced suppressive activity and impaired binding of GLI1. (A) Human rhabdomyosarcoma cells (CRL-2061) or (B) *Sufu* knockout (*Sufu*^{-/-}) mouse embryonic fibroblasts (MEFs) were transfected with wild-type or mutant *SUFU* constructs, Hedgehog pathway specific reporter plasmid and control reporter plasmids. The cells were incubated with and without ShhN-conditioned media and luciferase activity was measured. * t-test $p < 0.05$ compared to Myc-hSUFUWT. Error bars denote one standard deviation (duplicate samples). (C) The p.Arg123Cys mutant SUFU has reduced binding of GLI1. GLI1 binding in relation to Myc-SUFU protein level is shown below the panel.

4.5 Screening of *SUFU* in additional meningioma samples

To identify additional *SUFU* mutated patients, we Sanger sequenced the exons and exon-intron boundaries of *SUFU* from 162 meningioma-affected individuals, including 4 *NF2* mutation negative individuals with multiple meningiomas, 77 Finnish meningioma cases from the Interphone case-control study (Interphone Study Group, 2010), 35 individuals from the UK with multiple meningiomas, 5 familial UK meningioma cases, and 41 meningioma tumors collected through the Finnish Cancer Registry, but did not find any pathogenic mutations. Since *SUFU* is a known predisposition gene in childhood medulloblastoma (Taylor *et al.*, 2002), we also screened the c.367C>T mutation site from a distant relative of the Finnish meningioma family who had died of medulloblastoma at the age of 5. We screened the mutation site from DNA extracted from two different FFPE samples, but detected no *SUFU* mutations.

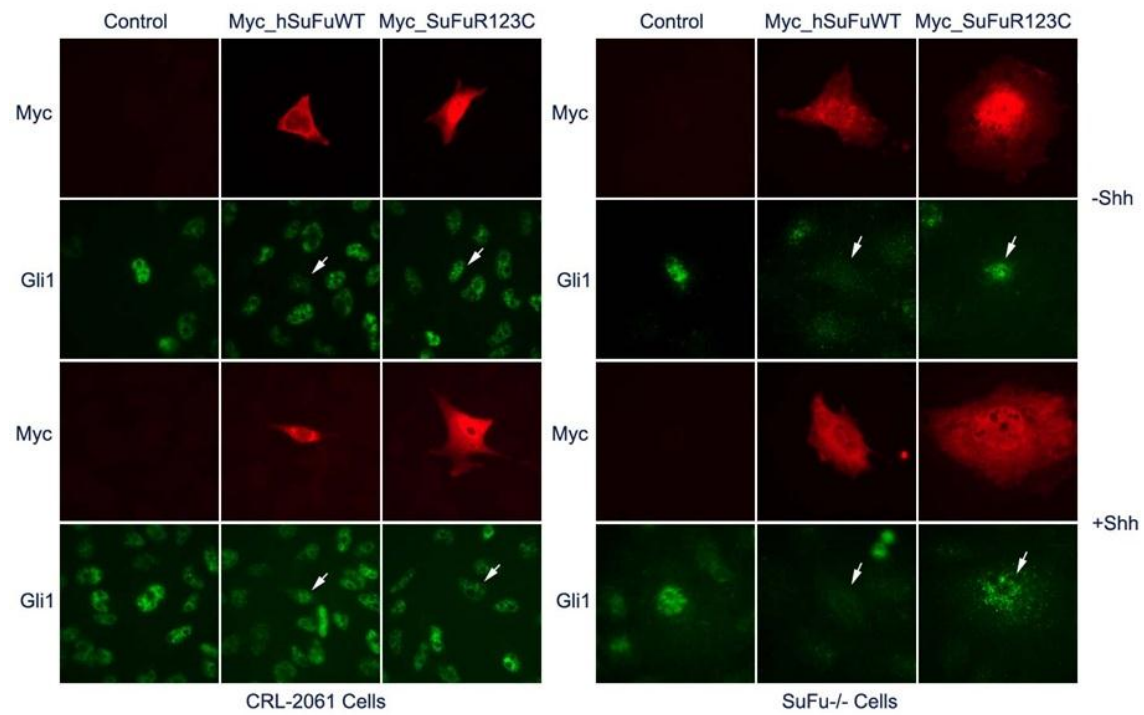


Figure 9. The p.Arg123Cys SUFU mutation impairs the re-localization of GLI1 from nucleus to cytoplasm in CRL-2061 and *SuFu* (-/-) mouse embryonic fibroblasts.

DISCUSSION

1. Familial aggregation of tumors in Finland

Cancer, like many other complex diseases, occurs mostly sporadically, but a small fraction of cases display familial inheritance. One aim of this study was to assess the familial aggregation of cancers in Finland. To meet this aim, we utilized the population-level information on Finnish inhabitants and cancer patients in the National Population Registry (NPR) and the Finnish Cancer Registry (FCR) databases, respectively. Both registries have high coverage and high quality data and use personal ID codes for unique identification, which enables linking of the data in the two registries (Pukkala, 2011).

NPR was established in 1969, and together with the local register offices it maintains and controls the data on Finnish citizens (<http://www.vrk.fi>). However, the bookkeeping of the inhabitants in Finland dates back to the 16th century and from 1660 onwards the Church Parish Registries have kept records of Finnish inhabitants on a national level, enabling extensive genealogy work and ancestry identification of individuals back to the 17th century. The ID code system in Finland was implemented in the 1960s and since 1967 a unique 11-digit ID code has been in use.

FCR is a population-level patient database, and it currently contains information on more than 1 million patients. FCR was founded in 1952 and it is one of the first nation-wide cancer registries in the world. The registration of incident cancers was made obligatory by law in 1961 and FCR covers almost all cancers in Finland since 1953 (Teppo *et al.*, 1994). In addition to cancer diagnoses, FCR collects information about the hospitals, laboratories and diagnostic samples of the patients. FCR is also an active research institute and produces the yearly cancer statistics in Finland (www.cancer.fi).

Because individuals with same family name and birth place are likely to be related, we used family name at birth and municipality at birth based clustering (MN-clustering) to assess the aggregation of cancers in Finland. To take into account the possible migratory events of the families we performed another clustering based on the tumor type and family name at birth only (N-clustering). Similar method has been employed by Matikainen *et al.* in 2000 to identify novel prostate cancer families (Matikainen *et al.*, 2000).

In contrast to many traditional epidemiological tools, our method can be used to study familial aggregation of both rare and common tumor types. The two systematic clustering efforts, that we performed yielded altogether 25 910 municipality-name based clusters and 12 695 name-based clusters, represented 183 different tumor types, which is considerably more than what previous studies on familial aggregation has been able to cover (Goldgar *et al.*, 1994; Amundadottir *et al.*, 2004; Albright *et al.*, 2012). The most commonly clustered tumor type was basal cell carcinoma of the skin, which was an expected result since it is the most common cancer in humans and the most registered tumor type in FCR.

When we ranked the different tumor types based on the cluster score, reflecting the frequency of cluster formation in relation to the incidence of a given tumor type in Finland, four out of five top

ranking tumor types represented tumors commonly seen in patients with hereditary cancer syndromes. The number one scoring tumor type was hemangioblastoma, which is a vascular tumor, most commonly affecting the nervous system. Hemangioblastomas occur frequently in patients with Von Hippel-Lindau syndrome (MIM #93300), caused by germline mutations in *VHL*. The second tumor type in our cluster score ranking list was medullary thyroid cancer. It is often seen in patients with multiple endocrine neoplasia 2A (MIM #171400) or 2B (MIM #162300) (Nose, 2011). The third most frequently clustered tumor type was KS, which is a soft tissue sarcoma with infectious etiology and unknown genetic predisposition (discussed in more detail in the next section). The fourth and fifth tumor types were pancreatic neuroendocrine tumor and nephroblastoma, indicative of patients with multiple endocrine neoplasia type 1 (MIM #131100) and Wilms tumors 1 (MIM #194070), respectively.

Neuroendocrine carcinoma of small intestine and myelosclerosis also showed strong familial aggregation in our cluster score assessment. Increased familial risk for small intestine neuroendocrine tumors has been previously reported (Hemminki and Li, 2001; Hiripi *et al.*, 2009; Kharazmi *et al.*, 2013), but the genetic etiology remains poorly studied. This is also the case with myelosclerosis. Landgren *et al.* described increased familial risk for almost all other myeloproliferative disorders, except for myelosclerosis. This however is most likely due to lack of statistical power in their study (Landgren *et al.*, 2008).

Strong familial clustering of squamous cell neoplasm and carcinoma of vulva and vagina was also detected in our data. Increased familial clustering of vulvar squamous cell carcinoma has also been previously reported (Hussain *et al.*, 2008). Human papillomavirus (HPV) is a causal factor in many anogenital cancers (IARC Working Group on the Evaluation of Carcinogenic Risks to Humans, 2012). The familial aggregation data by us and others, does not allow concluding whether the observed familial aggregation is caused by shared environmental or genetic factors. However in the case of vulvar squamous cell carcinoma, it is likely that the familial aggregation reflects factors related to HPV susceptibility, since HPV is mainly transmitted sexually.

From the more common tumor types, with more than 7000 patients in the FCR database, papillary thyroid adenocarcinoma, chronic lymphatic leukemia, and squamous cell carcinoma of the lip showed the strongest familial aggregation. Familial clustering of the same tumors has been previously observed in a study assessing the familial occurrence of cancers in the Utah Population Database, containing 85 550 cancer patients (Albright *et al.*, 2012). In addition, several other studies have described familial aggregation of papillary thyroid adenocarcinomas and reported up to 8-10 fold increased risk for relatives of papillary thyroid carcinoma patients (Nose, 2011; Bauer, 2013). A number of studies have demonstrated increased familial risk also for chronic lymphocytic leukemia (Gunz *et al.*, 1975; Giles *et al.*, 1984; Cartwright *et al.*, 1987; Pottern *et al.*, 1991; Goldgar *et al.*, 1994; Radovanovic *et al.*, 1994). Despite extensive efforts, no definite high risk susceptibility gene has thus far been identified for either one of these tumor types (Goldin *et al.*, 2010; Nose, 2011).

Since our clustering method has been developed to identify cases with similar disease, or other phenotypic feature, that share common birth place and birth name, it could in principle be used to

study familial aggregation of any other registered disease. The only requirement is that information on family names and birth places of the patients are available, while population-level data is required for statistical assessment of the clusters. Also, for the method to be able to identify statistically significant clusters, the family names in the study population should contain enough variation.

In Finland, the family names are generally inherited from father and thus our clustering method is restricted to identify siblings or families from the paternal side only. The method would heavily benefit from readily available genealogy data, like that available for example at the deCODE genetics in Iceland, the Swedish Family-Cancer Database in Sweden, or Utah Population Database in Utah, USA. A small pedigree, including the first degree relatives and children of siblings, could in principle be constructed based on the family data available in NPR, for almost all inhabitants in Finland. However, these data has limited coverage, covering only individuals born after 1967, or who were alive at that time, enabling the automatic pedigree formation only for a fraction of cancer cases in FCR.

1.1 Identification of potential tumor susceptibility families

This study also aimed to identify novel cancer families for subsequent research on genetic predisposition. As described above, the clustering data provides thousands of potential cancer families for further genetic analyses. KS, which produced significant MN-clusters third most commonly, was chosen as an example tumor type to study how well the method actually can predict true cancer families. We performed thorough genealogy work for all MN-clustered KS cases, resolving the family pedigrees at least three generations back. True relationships were confirmed for 70% of the individuals in MN-clusters. When we considered only the clusters with high statistical significance ($CI\text{-low} \geq 10$), 88% of the clustered cases were found to be related. Although it seems that the significance of the cluster can quite well predict true relation of the clustered individuals, a large amount of manual work is needed to be done to actually confirm them, and here again the presence of a nation-wide pedigree database would be an excellent resource.

The NLPHL family and the meningioma family examined in studies III and IV were originally observed by medical geneticists Kristiina Aittomäki and Pia Alhopuro, respectively. After thorough examination and genetic counselling, they were directed to us for further studies on genetic predisposition. This happened before study I was conducted, but in principle, both of the families could have been identified through our FCR clustering effort. Later examination of the clusters indeed revealed a MN-cluster of five meningioma patients ($p\text{-value } 0.06$ O/E 26 and $CI\text{-low } 188$) representing the meningioma family, while clusters representing the NLPHL family could not be detected. There are two possible reasons for this. Firstly, the mothers of the affected NLPHL cases had taken their spouses family names in the marriage, and thus the affected cousins did not share same family names. Secondly, the two affected brothers, who shared the family name, who should have formed MN- or N-cluster, had different lymphoma diagnoses (lymphoma, not otherwise specified [NOS] and Hodgkin disease) in the FCR, leading them to fall into two different tumor classes.

The NLPHL family example reflects the problematic and variable diagnosing of lymphomas, especially NLPHL, but it also reflects the difficulty of defining meaningful tumor classes for the clustering. The ICD-O-3 classification, currently in use in the FCR, is very detailed and to enable efficient cluster formation we had to re-classify the tumors into larger morphology and topography groups. Although this was done carefully, respecting the molecular and pathological backgrounds of different ICD-O-3 classifications, certain tumor types such as lymphomas, meloproliferative disorders, or hematological diseases might have benefited from less stringent grouping.

2. Kaposi sarcoma in the Finnish population

Although KS is a rare tumor type in most places in the world, it is the most common tumor in HIV-infected individuals and in certain parts of Sub-Saharan Africa (Mesri *et al.*, 2010). HHV8 is the infectious agent underlying the development of KS, and it is well acknowledged that patients suffering from immunosuppressive conditions, or lowered immunity due to immunosuppressive drugs, have higher risk for KS (Beral *et al.*, 1990; Mbulaiteye and Engels, 2006). To date, a little over 100 KS families, including the families identified in study I, have been described (Almohideb *et al.*, 2013), and accumulating evidence suggests that host genetic factors play an important role in KS development.

In study I, we observed a significant familial clustering of KS in Finland. KS is a very rare tumor type in Finland, with an average incidence of 0.1–0.2 per 100 000 person-years in 1963–2010, age-adjusted to the World Standard Population. We showed that the incidence of KS varied strongly according to the municipality of birth of the patients, with much less KS cases originating from eastern than western Finland. Unequal geographical distribution of KS has also been reported in Sweden, where the incidence of KS has been observed to be higher in northern parts of Sweden compared to southern parts (Hjalgrim *et al.*, 1996).

The prevalence of HHV8 varies strongly in different populations, and there seems to be a positive correlation between HHV8 prevalence and KS incidence (Tedeschi *et al.*, 2006, Mesri *et al.*, 2010). Thus, we hypothesized that the geographical difference in KS incidence in Finland could be due to uneven distribution of HHV8 infection. We selected 200 healthy age- and sex-matched individuals from Eastern and Western Finland and studied the prevalence of HHV8 in their sera. The overall prevalence of HHV8 was shown to be 1%, with two seropositive individuals originating from eastern and two from western Finland indicating similar level of HHV8 seropositivity. However, the result is far from definitive and a thorough examination of HHV8 seroprevalence in Finland would be important.

Molecular and serological data suggest that transmission of HHV8 occurs primary via saliva (Mayama *et al.*, 1998; Vieira *et al.*, 1997; Vitale *et al.*, 2000; Dupuy *et al.*, 2009) and that the risk of HHV8 infection is increased in first degree relatives of classic KS patients (Guttman-Yassky *et al.*, 2005; Mancuso *et al.*, 2011). While some of the KS families with two affected individuals may be explained by sharing of the HHV8 infection, it is highly unlikely that HHV8 alone could explain the KS in a family with as many as five affected cases.

2.1. *STAT4* and Kaposi sarcoma

In study II, we concentrated to identify genetic predisposition in the family of five affected individuals with classic KS. We mapped the shared chromosomal regions from two affected family members and combined the regions with variants identified from the genomes and exome sequenced from three affected family members. Because of the viral etiology of KS, we chose to use less stringent filtering criteria for the genetic variants and used a MAF cut-off of 0.005% in the control genomes (n=242). This led us to identify 14 candidate variants that were present in all 3 KS affected individuals in the family from whom we had samples.

Despite many efforts, we were not able to derive samples from all the affected family members, and thus the segregation of the 14 candidate variants could not be thoroughly examined. The mother of the three affected siblings had died in 1962 and her diagnostic blocks had already been destroyed. She, however, is an obligate carrier of all the 14 variants. We were also unable to derive sample from one of the affected siblings, because her diagnostic blocks were missing. Due to the fact that we were not able to study the segregation of all the candidate variants completely, we had to rely on more candidate-gene based approach, and we chose to concentrate on candidates residing in genes with human immunity related functions including *HLA-C*, *HLA-DRB5*, and *STAT4*.

Human leukocyte antigen (HLA) complex alleles have been suggested to be relevant for KS predisposition. Especially the HLA-DR5 serotype has been associated with increased risk for KS (Papasteriades *et al.*, 1984; Iannidis 1995). The HLA region of the genome is extremely challenging to study with the available sequencing methods, because of the high homology of HLA genes, and the high variability within the HLA region. Two of our human immunity related candidate variants resided in HLA genes. Closer inspection of the variants revealed that the c.106G>A in *HLA-C* is in fact relatively frequent in the population and that the *HLA-DRB5* variant was not predicted damaging for the protein function. This left the c.1337C>T (p.Thr446Ile) in *STAT4* to remain as the most prominent human immunity related variant in the KS family. The *STAT4* c.1337C>T (p.Thr446Ile) variant was *in silico* predicted to harm the correct tertiary structure of STAT4 and it was absent in the additional 360 Finnish control chromosomes examined.

We studied *STAT4* from a relatively large number of familial and sporadic KS cases, but apart from a single synonymous change (c.1338C>A, p.Thr446Thr; rs144421302), present in two sporadic KS cases, we did not detect any protein code altering variants. We also studied seven Finnish familial KS cases for mutations in *STAT4* associated genes and identified a novel, potentially damaging, heterozygous variant in *C-C chemokine receptor type 5* (*CCR5*, c.866A>G, p.His289Arg). *CCR5* encodes a chemokine receptor that also acts as a co-receptor of HIV-1 entry (Deng *et al.*, 1996; Dragic *et al.*, 1996). Individuals with specific *CCR5* mutations are reported to be resistant to HIV-1 infection (Dean *et al.*, 1996; Liu *et al.*, 1996; Samson *et al.*, 1996; Kaur *et al.*, 2013). The interaction of STAT4 and CCR5 is not well understood, but STAT4 has been described to be a negative regulator of *CCR5* expression in mouse T cells (Iwasaki *et al.*, 2001; Watford *et al.*, 2003).

STAT4 is located at 2q32.3 and encodes a transcription factor belonging to a family of Signal Transducer and Activator of Transcription factors including, *STAT1*, *STAT2*, *STAT3*, *STAT5A*, *STAT5B*, and *STAT6*. *STAT4* is expressed in the sperm cells, myeloid cells, and T lymphocytes and it is activated in response to certain growth factors and cytokines, especially interleukin-12, interleukin-23, and type 1 interferons (Takeda and Akira, 2000). Upon activation, *STAT4* forms dimers and relocates to the nucleus to regulate the transcription of its target genes, especially IFN- γ (Zhang and Boothby, 2006; Good *et al.*, 2009). We studied the IFN- γ production in the p.Thr446Ile mutation carriers and matched wild-type controls and detected statistically significant reduction in IFN- γ levels in the PMA-ionomycin induced naive T helper cells of the mutation carriers. Similar trend was also present in the memory T helper cells of the mutation carriers. We also studied if the mutation affected the levels of the phosphorylated *STAT4*, but these remained at comparable levels in the mutant carriers and the wild-type individuals.

The p.Thr446Ile disrupts an evolutionarily highly conserved threonine located in the DNA-binding domain of *STAT4*. The DNA binding domains of all STATs, but especially *STAT1* and *STAT4*, are highly homologous with over ~60% similarity. Germline mutations in IL-12 receptor gene *IL-12R β 1* and *STAT1* cause Mendelian susceptibility to mycobacterial disease (MIM #209950) (Filipe-Santos *et al.*, 2006). Specific alleles (p.Glu463H and p.ProE320Glu) affecting the DNA-binding domain of *STAT1*, have been reported in otherwise healthy patients with mycobacterial disease (Chapier *et al.*, 2006). Thus it is interesting to speculate whether the p.Thr446Ile *STAT4* mutation might also be responsible for the index case's chronic atypical mycobacterial infection.

Stat4 knock-out mice are viable but have severe problems in the IL-12 mediated functions, including T helper 1 cell differentiation, IFN- γ production, cell proliferation, and activation of natural killer cell cytolytic functions (Kaplan *et al.*, 1996; Thierfelder *et al.*, 1996). Moreover, they are susceptible to various intracellular bacteria and pathogens (Kaplan, 2005). A recent forward genetics screen discovered a *Stat4* DNA-binding domain mutation (p.Gly418_Glu445del) in mice susceptible for Acute Typhoid-like disease (Eva *et al.*, 2014). Already the heterozygous mutant mice had defective IFN- γ mediated immunity, comparable to our cKS family members (Eva *et al.*, 2014). In addition to this, a common *STAT4* haplotype has been associated with susceptibility to certain autoimmune diseases, including rheumatoid arthritis, systemic lupus erythematosus and primary Sjögren's syndrome (Remmers *et al.*, 2007; Lessard *et al.*, 2013). The common associated SNP (rs7574865) located within the haplotype has also been associated with risk of hepatitis B virus-related hepatocellular carcinoma (Jiang *et al.*, 2013). One report indicates *STAT4* to regulate the antiviral IFN- γ responses during genital herpes simplex virus 2 infection (Svensson *et al.*, 2012). Intriguingly, the p.Thr446Ile *STAT4* allele has been detected previously from one Finnish individual with high serum IgE levels (Pykäläinen *et al.*, 2005).

As a summary, we identified 14 possible KS predisposing candidate variants, among them a putatively damaging heterozygous missense variant located in the DNA-binding domain of *STAT4*. We provided preliminary evidence that the c.1337C>T *STAT4* mutation carriers may have defective IFN- γ mediated immunity due to reduced levels of IFN- γ produced by their activated CD4+ T cells, but we were not able to genetically validate *STAT4* in other KS patients.

Based on our data, *STAT4* remains as a candidate predisposition gene, and future studies are necessary to fully elucidate its role in KS predisposition.

3. NPAT in Hodgkin lymphoma predisposition

NLPHL is a rare HL subtype with distinct morphology, immunophenotype, and clinical features (Harris *et al.*, 1994). Not much is known about its etiology, and it is a challenging disease to study because the patients are so few and because the tumor material is difficult to obtain, since the malignant cells in the tumor constitute less than 1% of the cells. High increased risk in the first degree relatives of NLPHL patients has recently been reported (Saarinen *et al.*, 2013), and to date only five families with NLPHL have been described, including the family examined in study III (Saarinen *et al.*, 2011), two families with affected siblings (Campbell *et al.*, 2004; Ur Rehman *et al.*, 2008), one family with affected mother and son (Unal *et al.*, 2005), and a family with two affected brothers and father (Bauduer *et al.*, 2005).

To gain insight to the NPLHL predisposition, we performed linkage analysis and exome sequencing in the Finnish family of four NLPHL affected cousins. As the only plausible candidate, a heterozygous frameshift mutation (c.2437-2438delAG) in *nuclear protein, ataxia-telangiectasia locus* (*NPAT*) was identified. The mutation segregated with the NLPHL in the family and was absent in 478 control chromosomes sequenced from 239 healthy Finnish blood donors. In addition to the affected family members, the mutation was present in five other members of the family, including the oblique carrier mothers and two aunts, who were identical twins. The presence of healthy mutation carriers in such a family indicates incomplete penetrance, which is a common feature of hereditary cancer syndromes (Nagy *et al.*, 2004). After the publication of study III, however, we learned that one of the unaffected mutation carriers, the other twin aunt, had developed diffuse large B-cell lymphoma (DLBCL), which is the most common non-HL subtype. NLPHL and DLBCL both derive from B lymphocytes (Campo *et al.*, 2011), and a small fraction of NLPHLs are known to progress into DLBCL (Al-Mansour *et al.*, 2010; Biasoli *et al.*, 2010). Also, NLPHL is sometimes difficult to differentiate from a specific DLBCL subtype, T-cell/histiocyte-rich large B-cell lymphoma, and thus careful revision of the aunt's diagnosis might be worth conducting (Campo *et al.*, 2011). It would also be interesting to screen *NPAT*, from patients or tumor samples of DLBCL patients to see if it would also underlie the development of DLBCL.

To obtain genetic validation, we screened *NPAT* from a large set of samples (n=169), including NLPHL and HL patients from Finland and UK, as well as the NLPHL derived cell line (DEV). In addition to common polymorphisms, an in-frame deletion of serine 725 was identified. The deletion was identified altogether from seven patients and five controls. Statistical evaluation argued increased risk of HL with the variant (odds ratio 4.11), although this is likely to be an overestimation because of the relatively small sample size and enrichment of familial cases in our data. Nevertheless, the patients carrying the p.Ser725del in our data had either early onset of NLPHL or HL or had family history of lymphomas, indicating possible genetic predisposition. Interestingly, the same deletion is listed in the Catalogue of somatic mutation in cancer (COSMIC, <http://cancer.sanger.ac.uk/cosmic/>). COSMIC lists the p.Ser725del to be present in

cancer cell lines, originating from kidney (TK10), large intestine (KM12), and lung (NCI-H522) (Abaan *et al.*, 2013). As serines are often sites for post translational phosphorylation and since NPAT is known to be regulated by phosphorylation (Imai *et al.*, 1996; Ma *et al.*, 2000), it would be interesting to study whether the Ser725 would be a site of NPAT regulation.

We wanted to see if the c.2437-2438delAG mutation affected the expression of *NPAT*, and we performed transcriptome analysis which showed that the expression of *NPAT* was significantly reduced in the mutant carriers. This was also confirmed by cDNA sequencing. We saw a clear drop in the mutant allele intensities in all of the studied mutation carriers, suggesting that it was degraded by nonsense mediated decay pathway. Messenger RNAs (mRNAs) with premature termination codons are often directed to nonsense mediated decay pathway, which is a safety mechanism of the cells to destroy of non-functional mRNAs and reduce errors in gene expression (Baker and Parker, 2004; Maquat, 2005).

The great majority (about 90%) of high and moderate risk tumor predisposition genes are tumor suppressor genes harboring gene inactivating mutations (Rahman, 2014). The nature of the *NPAT* mutation, supported by its reduced expression, would thus suggest it to be a tumor suppressor gene rather than an oncogene. To confirm this, one could have for instance studied the presence of 'second hit' in the tumors of the mutation carriers. However, in the case of NLPHL, this is not possible to do by using bulk extracted tumor samples, because the lymphocyte-predominant cells, that are the tumor cells, constitute only about 1% of the cells in the tumor (Swerdlow *et al.*, 2008). We initially tried to microdissect the lymphocyte-predominant cells from the tumors of the affected individuals, but it proved to be extremely difficult to identify the lymphocyte-predominant cells without appropriate immunohistochemical and morphological stainings.

NPAT, also known as *EI14* or *CAND3*, is located very close to *ATM* in the major AT locus on human chromosome 11q22-q23 (Imai *et al.*, 1996). The two genes are transcribed to opposite directions and separated from each other only by 0.5kb (Imai *et al.*, 1996). Germline mutations in *ATM* cause ataxia-telangiectasia (MIM #208900), which is a recessively inherited disorder affecting the nervous and immune systems of the patients. Intriguingly, ataxia-telangiectasia patients are in increased risk of getting cancers, particularly B-cell lymphomas and T-cell leukemias (Garber & Offit 2005), making it intriguing to consider if the tumorigenic effect of *NPAT* might be associated with *ATM*.

Deletion of 11q22-q23, encompassing *NPAT* and *ATM* among other genes, is one of the most common genetic alterations in chronic lymphocytic leukemias, and the deletion has also been detected in many other B-cell malignancies (Coll-Mulet and Gil, 2009). Although *ATM* is usually considered to be the main tumor suppressor gene at the 11q22-q23 locus, biallelic mutations of *ATM* are not always detected, implicating presence of other tumor suppressor genes at the region. A study by Kalla *et al.* described reduced expression of *NPAT* in chronic lymphocytic leukemia patients with monoallelic deletion of 11q22-q23 (Kalla *et al.*, 2007). Another study by Mishra *et al.* reported promoter hypomethylation of *NPAT* in prostate cancer cell lines, further supporting its possible tumor suppressor function (Mishra *et al.*, 2010).

NPAT is abundant in most human tissues, and its expression peaks in the G1-S-phase transition of the cell cycle (Zhao *et al.*, 1998). *NPAT* is phosphorylated by the cyclin E-CDK2 complex, which also supports its central role in cell cycle progression (Zhao *et al.*, 1998). Overexpression of cyclin E is commonly observed in many tumors, including multiple lymphoma subtypes (Erlason 1998 Hwang & Clurman 2005). Activation of the cyclin E-CDK2 complex is controlled by the Myc and RB/E2F pathways, defected in most of the human cancers (Hanahan and Weinberg, 2000). In lymphomas, Myc pathway defects are frequently observed in Burkitt lymphoma (Smith *et al.*, 2010), and RB/E2F pathway defects in classical HL (Bai *et al.*, 2005). *NPAT* is also shown to regulate S-phase-specific histone gene transcription together with HINFP (Ma *et al.*, 2000; Xhao *et al.*, 2000). In addition to the S-phase-specific histone genes, the HINFP-*NPAT* complex has also been described to activate promoters of non-histone genes, including *NPAT* and *ATM* (Medina *et al.*, 2007).

As a summary, *NPAT* is an important regulator of the cell cycle, and based on the current knowledge it can be linked to the development of HL through a number of different mechanisms. We identified a truncating *NPAT* mutation in one NLPHL family, and although higher risk of HL was observed in individuals with the in-frame deletion of *NPAT* Ser725, the data is hardly strong enough to validate that the relevance of *NPAT* in HL. More NLPHL or HL patients, harboring germline or somatic *NPAT* mutations should be identified, and additional experiments conducted to confirm the possible function of *NPAT* in the development of HL.

4. SUFU, Hedgehog signaling, and meningioma

Meningiomas are the most common primary tumors affecting the brain and spinal cord (CBTRUS, 2012; Wiemels *et al.*, 2010). Although majority of meningiomas occur in isolated cases, a small fraction of meningiomas show familial accumulation. A fraction of familial meningiomas arise in individuals with Neurofibromatosis type II (MIM #101000), which is an autosomal dominant disorder characterized by central and peripheral neural system tumors, especially acoustic neuromas. A smaller subset appears in patients with familial schwannomatosis (MIM #162091), which is an autosomal dominant tumor susceptibility condition characterized by cutaneous neurilemmomas and central nervous system schwannomas. More recently, a third group of patients, with familial clear cell meningiomas, were identified to harbor germline mutations in *SMARCE1* (Smith *et al.*, 2013). However, a small group of familial meningioma patients, with intracranial multiple tumors, describe a distinct entity without association to the abovementioned conditions.

In study IV, we examined a Finnish family of five siblings with meningiomas, four of whom had multiple tumors. To identify the genetic defect, we conducted linkage analysis and sequenced the germline exomes of three affected family members. After combining the two data sets, 20 protein code altering candidate variants remained, of which 7 were further shown to segregate in all the 5 affected meningioma patients. The presence of these seven variants was further examined in the healthy family members and in additional population-based controls. Two of the variants were excluded from further consideration, because they were present in unaffected siblings and three were excluded because they were present in the population-based controls, leaving two possible

candidate variants, namely the c.367C>T (p.Arg123Cys) in *SUFU* and the c.1990A>G (p.Ser664Pro) in *AASS*. From these the c.367C>T (p.Arg123Cys) in *SUFU* was *in silico* predicted damaging, while the c.1990A>G (p.Ser664Pro) in *AASS* was predicted benign. Also, no LOH was observed at the site of the *AASS* variant, whereas at the site of *SUFU* variant LOH was observed in each of the studied tumors, strongly indicating a presence of a tumor suppressor gene.

Suppressor of Fused Homolog (Drosophila) (SUFU) is located at 10q24.32, and it encodes a key negative regulator of the Hedgehog signaling (Hh) pathway. The pathway is essential for correct embryonic development and it is present in all bilateral animals (Ingham *et al.*, 2011; Varjosalo and Taipale, 2008). Hh cascade is triggered by binding of the ligand to the receptor (PTCH1 or PTCH2). This interaction leads to a release of SMO and activation of GLI family of transcription factors. The activated GLIs moves to the nucleus to regulate expression of the Hh pathway target genes (Varjosalo and Taipale, 2008; Ingham *et al.*, 2011). Hh pathway is controlled by many molecules, but in humans *SUFU* is the main negative regulator (Svard *et al.*, 2006; Varjosalo *et al.*, 2006; Ruel and Therond, 2009). In addition to the central role of Hh pathway in embryogenesis, it is also important in the adult tissues, where it regulates the proliferation of various stem and progenitor cells, including neural progenitor cells (Lai *et al.*, 2003; Varjosalo and Taipale, 2008).

SUFU germline mutations predispose to medulloblastomas, which are aggressive brain tumors affecting generally young children (Taylor *et al.*, 2002; Cheng and Yue, 2008). Most of the medulloblastoma predisposing *SUFU* mutations are truncating, and the tumors commonly also display a ‘second hit’. Somatic inactivating *SUFU* mutations are also often observed in medulloblastomas and basal cell carcinomas of the skin (Taylor *et al.*, 2002). The *SUFU* mutation that we identified was a missense change, and thus the effect of the mutation in principle could be *SUFU* activating. However, the presence of LOH at the site of the *SUFU* mutation in the studied meningioma tumors strongly implies that also in the case of meningiomas *SUFU* acts like a tumor suppressor gene. This was further confirmed by our *in vitro* experiments that demonstrated that the p.Arg123Cys mutation impairs the binding of *SUFU* with GLI1, disturbs the *SUFU* mediated nuclear-cytoplasmic shuttling of GLI1 (Kogerman *et al.*, 1999; Shi *et al.*, 2014), and significantly reduces its ability to suppress Hh pathway. Although the p.Arg123Cys mutation is clearly damaging for the correct functioning of *SUFU*, it does not totally abolish its function, and it is interesting to speculate that the remaining activity may explain the phenotypic difference between meningiomas and medulloblastomas.

A large chromosomal deletion encompassing the long arm of chromosome 10 is also frequently seen in sporadic meningiomas (Louis *et al.*, 2007). The 10q deletion has been reported to affect especially anaplastic, atypical and malignant meningiomas (Louis *et al.*, 2007). The deleted region is usually large and involves multiple potential tumor suppressor genes. Although *PTEN* is often considered to be the main tumor suppressor gene at this region, a study comparing the 10q deletions in glioblastomas and meningiomas reported biallelic *PTEN* mutations in a subset of gliomas but not in meningiomas (Bostrom *et al.*, 1998).

In addition to medulloblastoma predisposing germline *SUFU* mutations, germline mutations in *SUFU* have been identified from a family and an isolated case with basal cell nevus syndrome

(MIM #109400) (Pastorino *et al.*, 2009; Kijima *et al.*, 2012). Basal cell nevus syndrome, also known as Gorlin syndrome, is an autosomal dominant disorder characterized by multiple basal cell carcinomas, odontogenic keratocysts of the jaws, and skeletal abnormalities (Lo Muzio, 2008). Basal cell nevus syndrome is usually caused by inherited mutations in the Hh pathway receptor *PTCH1* (Lo Muzio, 2008). In addition to basal cell carcinomas, the basal cell nevus syndrome patients have increased risk of medulloblastoma (Lo Muzio, 2008). Although meningiomas are not generally considered to belong to the tumor spectrum, some reports show meningiomas in patients with basal cell nevus syndrome (Tate *et al.*, 2003; Sobota *et al.*, 2007; Pribila *et al.*, 2008; Kijima *et al.*, 2012). Intriguingly, the isolated basal cell carcinoma patient with germline *SUFU* mutation, described by Kijima *et al.* had developed a meningioma (Kijima *et al.*, 2012).

Recent high-throughput sequencing studies on meningiomas have identified *SMO* mutations in about 5% of sporadic meningiomas (Brastianos *et al.*, 2013; Clark *et al.*, 2013). These tumors display a distinct meningioma entity, characterized by activated Hh pathway (Brastianos *et al.*, 2013; Clark *et al.*, 2013). The *SMO* mutations identified in meningiomas thus far hit the same mutational hotspot often mutated in basal cell carcinomas and medulloblastomas (Reifenberger *et al.*, 1998; Jones *et al.*, 2012). Multiple *SMO* inhibitors are currently in clinical development and Vismodegib (GDC0449, Curis/Roche) has recently been approved by the US Food and Drug Administration, for the treatment of basal cell carcinomas (Rudin, 2012; Amakye *et al.*, 2013). Although majority of meningiomas are treated surgically, some are inoperable and many recur after successful operation. Thus it would be important to study if a fraction of meningioma patients would benefit from treatment with *SMO* inhibitors.

We identified a germline mutation in *SUFU* that segregates with the meningiomas in a family of five affected siblings. The mutation was not found from the population based healthy controls, and all the studied tumors displayed LOH at the site of the mutation. We also showed that the *SUFU* mutation impaired the interaction with *GLI1* and led to dysregulated Hh pathway. This data, together with the recent discovery of *SMO* mutations in sporadic meningiomas, indicate that *SUFU* is a meningioma predisposition gene, and that dysregulated Hh pathway underlie the development of a small but significant subset of meningiomas.

CONCLUDING REMARKS AND FUTURE PROSPECTS

Cancer is one of the leading causes of disease and death world-wide, with an estimated number of 14.1 million new cancer cases and 8 million cancer related deaths world-wide during 2012 (Ferlay *et al.*, 2013). Solely in Finland, 30 314 new cancers and 11 860 cancer related deaths were registered in 2012 (<http://www.cancer.fi/syoparekisteri/en/statistics/>).

Majority of tumors occur due to environmental causes, and only a small portion occurs due to inherited defective cancer susceptibility genes. Many cancers also display familial aggregation without clear Mendelian inheritance. Various excellent molecular genetic techniques have been developed and widely employed during the past few decades to understand the genetic susceptibility to cancer, and a number of great discoveries have been made, especially in the field of hereditary cancer syndromes and Mendelian tumor susceptibility genes. It is sometimes claimed that most of the genes underlying hereditary cancer syndromes have already been identified. However, during the last five years, multiple novel high penetrance cancer susceptibility genes have been identified with the help of high-throughput sequencing methods (Rahman, 2014).

This study was conducted to assess the familial aggregation of tumors in Finland, to identify novel tumor susceptibility families and to elucidate the genetic predisposition of familial NLPHL, meningioma, and KS. The conclusions are as follows:

I) By using family name at birth and municipality at birth as identifiers for familiarity, we performed a nation-wide assessment of familial aggregation of 183 tumor types in Finland. As a result we identified 25 910 birth name-municipality and tumor type based clusters and 12 695 birth name-tumor type based clusters. We evaluated the familial occurrence of the tumor types and identified frequent familial clustering of KS in Finland. Furthermore, the clustered as well as the sporadic KS cases were shown to accumulate to the western parts of Finland, suggesting shared genetic or environmental factors underlying KS. We also identified a family with five affected individuals in two generations and several smaller families with two first degree relatives with KS. This work showed that the birth municipality- and birth name- based clustering is a powerful method to identify familial cancer cases, and future studies will hopefully elucidate the genetic basis of other potential predisposition conditions found in this study.

II) To elucidate the underlying genetic factors behind classical KS, we mapped the shared chromosomal regions and sequenced two genomes and an exome from three affected individuals belonging to the Finnish family of five affected individuals. We identified 14 potential KS predisposing candidate variants, among them a rare missense variant c.1337C>T (p.Thr446Ile) in the DNA binding domain of *STAT4*. We showed that the healthy mutant carriers had decreased IFN- γ production compared to the matched wild-type controls. We conducted a search for additional *STAT4* mutations in a large set of KS cases, but found no pathogenic changes. We also studied the *STAT4* related genes in a set of seven familial KS cases and identified a putatively harmful missense change in *CCR5* that was absent in the studied controls. However, further genetic and functional studies are needed to draw conclusions of the importance of the *STAT4* and the related genes in KS predisposition.

III) To study the genetic predisposition in a rare HL subtype, NLPHL, we combined genome-wide linkage analysis and exome sequencing. A frameshift mutation in *NPAT* was described and found to segregate with NLPHL in a Finnish family with four affected cousins. Additionally, an in-frame deletion of serine 725 of the *NPAT* was detected, and it was shown to be more prevalent in the HL cases than in the healthy controls supporting a role of *NPAT* in the HL predisposition. However, further genetic and functional studies confirming the relevance of *NPAT* in the genesis of HL should be performed.

IV) Linkage analysis and exome sequencing was utilized to identify a c.367C>T (p.Arg123Cys) mutation in *SUFU* underlying the intracranial meningiomas in a Finnish family of five affected siblings. Loss-of-heterozygosity of *SUFU* was observed in all the studied tumors, indicating a classic tumor suppressor gene function for *SUFU*. *In vitro* and *in silico* experiments were conducted, and the *SUFU* mutation was shown to affect its function and lead to constitutively activated Hh pathway. We studied a large number of meningioma cases to identify additional *SUFU* mutations, but were unable to detect any mutations that would have been pathogenic. Recent studies in sporadic meningioma have revealed Hh pathway dysfunction in approximately 5% of meningiomas (Brastianos *et al.*, 2013; Clark *et al.*, 2013). Larger studies revealing the extent of Hh pathway dysregulation in meningiomas are needed.

The few past few years have witnessed a revolution in the field of human genetics. The sequencing of the first haploid human genome took over ten years and an incredible amount of work and effort from hundreds of scientists. Today it is feasible to sequence a genome of an individual within one working day. The development of novel sequencing methods has not ceased, and new techniques are constantly developed. Little by little, as their use has become more routine, high-throughput sequencing based methods are entering clinical use, and hopefully in the near future the era of actual personalized medicine will become reality.

As any new technology, the high-throughput sequencing methods also provide many new challenges. Increasing amount of rare variants, both in patients and the general population are detected. Determining the true disease related variants from individual rare variation is not trivial, especially in the cases of very rare conditions, where it is difficult to find supportive genetic evidence. A rigorous validation, both functional and genetic, will be needed, and causative variants are to be “handled with care”, that is, to be claimed causative only with solid evidence. Most likely international collaboration and meta-analytical approaches will help in this.

Although the new techniques have opened up possibilities to almost any scientist to zoom in to the base-pair level of individual genomes, a utopist dream that a geneticist might have had a few decades ago, we still have only a vague understanding of the causes of familial clustering of cancers outside the framework of monogenic inheritance. To be able to tackle the etiology of familial aggregation outside monogenic inheritance, increased understanding of the regulatory genome and gene-environment interplay will be needed.

Finland has long traditions in successful genetic research. Many factors have contributed to this, including comprehensive public health care system, high quality nation-wide registry infrastructure, as well as the genetic background of the Finnish population. Important aspects

include also the willingness of the Finns to participate in genetic studies and the Finnish legislation, which generally supports registry-based studies. This thesis work utilized the possibility to conduct a registry-based population-level study on familial aggregation of tumors and focused on the identification of novel high and moderate risk genetic alterations from Finnish families with rare and common tumor types. The work would not have been possible without the willingness of the families to participate in the study, or without the excellent data at the FCR, not forgetting the Finnish legislation supporting the conduction of these kinds of studies, and it is certainly crucial that the situation remains the same in the future.

ACKNOWLEDGEMENTS

This work was carried out in 2009-2014 in the Department of Medical Genetics, Haarman Institute, and Genome-Scale Biology Program, Research Programs Unit, Biomedicum Helsinki, Faculty of Medicine, University of Helsinki. I wish to thank the present and the former heads of the departments for providing inspiring working environment and excellent research facilities.

Helsinki Graduate Program in Biotechnology and Molecular Biology (GPBM) and Integrative Life Science graduate program (ILS) are acknowledged for offering high quality graduate education as well as financial support. This thesis was also supported by Biomedicum Helsinki Foundation, the Finnish Cancer Society, the Ida Montini Foundation, the Maud Kuistila Memorial Foundation, the Orion-Farmos Research Foundation, and the Paulo Foundation, all of which are gratefully acknowledged.

I am greatly indebted to my supervisor Lauri Aaltonen for his immense support and trust in me during these years. Lauri's almost fatherly style of leading a large research group, and conducting outstanding medical genetic research, while still caring about its every member, is second to none. Above all, I would like to express my gratitude to Lauri for always taking me as who I am. I am also deeply grateful to my second supervisor Pia Vahteristo, who has always been extremely kind, encouraging and constructive towards me and my work. I admire Pia's attitude towards science and teaching, and I can only hope that some of that will stay with me in the future.

Minna Männikkö and Samuli Ripatti are thanked for reviewing the thesis book and giving valuable comments to improve it. Tiia Pelkonen is also acknowledged for the language review of the thesis book. Samuli Ripatti (again) and Laura Sarantaus are acknowledged for being part of my thesis committee.

I wish to express my sincere gratitude to all collaborators and co-authors. Kristiina Aittomäki, Pia Alhopuro and Minna Pöyhönen are thanked for their efforts and help regarding the meningioma and NLPH-families. Ari Ristimäki and Kaarle Franssila are acknowledged for expertise in pathology. Eero Pukkala, Miia Artama, Maarit Laaksonen, Toni Patama, Paul Knekt and Hilikka Laasanen are thanked for the collaboration and expertise in the Finnish Cancer Registry related matters. Kristian Ovaska is thanked for effortless, professional and nice collaboration regarding the TumorGene-project. Outi Vaarala, Janne Nieminen, Sinikka Tsupari and Harri Salo are warmly thanked for the help and input in the Kaposi sarcoma project, as are also Päivi Ojala and Pirita Pekkonen. Song-Ping Li is thanked for all the functional work with SUFU and Jussi Taipale for expertise in Hedgehog signalling. Ekaterina Morgunova is thanked for the help in the protein modelling in meningioma and Kaposi sarcoma projects and Minna Taipale for the high-throughput sequencing services. Co-authors Anssi Auvinen, Barbara Baim, Frederic Bauduer, Peter Broderic, Minji Buyin, Jean-Laurent Casanova, Rosie Cooke, Gareth Evans, Pasquale Ferrante, Richard Houlston, Anne Kiuru, Virpi Launonen, Roberta Mancuso, Markus Mäkinen, Ronit Sarid, Miriam Smith, Anthony Swerdlow, Jukka Tarkkanen, and Ali Ünal are also sincerely acknowledged for invaluable help and discussions.

My warmest thanks go to all the members of the Aaltonen lab. It has truly been fun to work with all of you!

Big thanks go to Sini N, Sirpa, and Marjo, who have done a great deal of work for this study. I am sincerely thankful for the extremely professional help and detective work regarding various sample collections, genealogy work, and numerous other things. Henna and Minna are also greatly thanked for expertise in genealogy work. Sini N is specially thanked for arranging BBQ-parties and also for excellent company in the sample collection trips in the spring 2013. Warmhearted thanks belong also to Alison, Inkku, Iina V, Mairi, Maarit, Sini K and Heikki M. Thank you for your company and for the numerous occasions that you have helped me! A special thanks to Alison for language help!

My officemates (=the Incubators) Eevi, Elina, Heikki M, Heikki R, Iikki, Johanna, Manuel, Miika, Riku, and Yilong. We may have shared a small, noisy, and crowded office, but because of you it has always felt like a home! Thanks for all the laughs and the most random discussions (also the few scientific ones! ;). I want to thank Riku for all bioinfo help, Mazda-partnership, teaching me how to snowboard, and also for getting Heikki M to join the Aaltonen lab... A special thank goes also to Iikki, who has kept my blood sugar levels high the last few years. I would also like to thank Miika for many interesting discussions, and for the company in many trips and parties. Johanna is also thanked for company and friendship. I am grateful for your help in numerous situations in the life where I have needed a 'doctor's opinion'.

A special thank goes to Eevi, who has been my closest collaborator and the person in the lab with whom I have shared the most projects and scientific discussions with. Thanks for the countless times you have pushed me to do things and finish off projects. Your uncompromising attitude towards work is deeply admirable, and I like to think that your example has taught me to aim a little higher. I am also deeply indebted for your friendship. With you the time at Aaltonen lab has truly been an interesting trip - and I hope there are many more still to come!

I also want to thank my colleagues Anniina, Hande, Hanna, Iina T, Jaana, Kati, Netta, Silva, Taru, Tati, Tomas, and Ulrika. Silva is especially thanked for very efficient and professional collaboration in various projects, and also for great relaxing times outside work! Kati is dearly thanked for company on therapeutic horse riding trips and Netta is thanked for interesting conversations about life and science in our unofficial gettogethers with Johanna, Silva and Eevi. Tati, thanks for hosting an unforgettable trip to Menorca and teaching us how to Fiesta!

Post docs: Auli, Alex, Esa, Heli L, Javier, Kimmo, Linda, Niko, Outi, Rainer, Sari, and Tuomas are greatly acknowledged for all the help, mentoring, and encouragement. Auli you are a unique personality, whom I have had pleasure to get to know and whose unselfish help I have been able to count on, in so many different scientific and non-scientific questions. Alex and Heli, you have been an essential part of my time in the Aaltonen lab and I want to deeply thank you for friendship and all sorts of fun events and trips that we have had together. Alex is also thanked for unforgettable and crazy times organizing ScienceSLAM Helsinki.

Sofia, Saija, Jenny, Denis, and Maral are thanked for friendship and above all peer support! Friends outside Biomedicum, especially Anna, Hannu, Inka, Jenni Luolaja, Jenni Mellavuo, Kirsi, Laura, Linda Koivunen, Linda Kolmonen, Ninni, Pekka, Pia, Sara, and Soila. Thanks for all the support and for reminding me about the truly important matters of life. You are invaluable to me!

I am deeply grateful to my family: my parents Ilkka and Saara and my siblings Ellu, Reino, and Veera, and the youngest of our family, Frida and Laila. Ilkka and Saara, thank you for always being so kind, supportive, and understanding. Reino, Ellu and Veera thanks for being there, your examples have influenced me probably much more than you can ever imagine. Also, I want to kindly thank Heikki (M) for his love and support. Thanks for showing me how powerful it can feel to care and be cared in return.

Finally, I would like to direct my sincere appreciation to all the individuals and families who have participated in this study.

Helsinki, September 2014

Mervi Aavikko

WEB-BASED RESOURCES

1000 Genomes	http://browser.1000genomes.org/
Brainarray	http://brainarray.mbnl.med.umich.edu/brainarray/
Chromas	http://technelysium.com.au/?page_id=13
Clustal W	http://www.ebi.ac.uk/Tools/msa/clustalw2/
COSMIC	http://cancer.sanger.ac.uk/cancergenome/projects/cosmic/
dbSNP	http://www.ncbi.nlm.nih.gov/SNP/
Ensembl	http://www.ensembl.org/
ExonPrimer	http://ihg.gsf.de/ihg/ExonPrimer.html
Exome Variant Server	http://evs.gs.washington.edu/EVS/
FastQC	http://www.bioinformatics.babraham.ac.uk/projects/fastqc/
Finnish Cancer Registry	http://www.cancer.fi/syoparekisteri/en/
Genome Reference Consortium	http://www.ncbi.nlm.nih.gov/projects/genome/assembly/grc/
ICGC	http://dcc.icgc.org/
NCBI	http://www.ncbi.nlm.nih.gov/
OMIM	http://www.omim.org/
National Population Registry	http://www.vrk.fi/
Protein data bank	http://www.rcsb.org/pdb/home/home.do
Picard Tools Markduplicates	http://picard.sourceforge.net/command-line-overview.shtml
Polyphen2	http://genetics.bwh.harvard.edu/pph2/
Primer3	http://bioinfo.ut.ee/primer3-0.4.0/
R	http://www.r-project.org/
SIFT	http://sift.jcvi.org/
Swiss-Model	http://swissmodel.expasy.org/
UCSC	http://genome.ucsc.edu/
Uniprot	http://www.uniprot.org/

REFERENCES

- CBTRUS Statistical Report: Primary Brain and Central Nervous System Tumors Diagnosed in the United States in 2004-2008. 2012. 1-58. Central Brain Tumor Registry of the United States, www.cbtrus.org
- 1000 Genomes Project Consortium, Abecasis, G. R., Auton, A., Brooks, L. D., DePristo, M. A., Durbin, R. M., Handsaker, R. E., Kang, H. M., Marth, G. T. & McVean, G. A. 2012. An integrated map of genetic variation from 1,092 human genomes. *Nature* 491: 56-65.
- Abaan, O. D., Polley, E. C., Davis, S. R., Zhu, Y. J., Bilke, S., Walker, R. L., Pineda, M., Gindin, Y., Jiang, Y., Reinhold, W. C., Holbeck, S. L., Simon, R. M., Doroshow, J. H., Pommier, Y. & Meltzer, P. S. 2013. The exomes of the NCI-60 panel: a genomic resource for cancer biology and systems pharmacology. *Cancer research* 73: 4372-4382.
- Abecasis, G. R., Cherny, S. S., Cookson, W. O. & Cardon, L. R. 2002. Merlin--rapid analysis of dense genetic maps using sparse gene flow trees. *Nature genetics* 30: 97-101.
- Adzhubei, I. A., Schmidt, S., Peshkin, L., Ramensky, V. E., Gerasimova, A., Bork, P., Kondrashov, A. S. & Sunyaev, S. R. 2010. A method and server for predicting damaging missense mutations. *Nature methods* 7: 248-249.
- Albright, F., Teerlink, C., Werner, T. L. & Cannon-Albright, L. A. 2012. Significant evidence for a heritable contribution to cancer predisposition: a review of cancer familiality by site. *BMC cancer* 12: 138-2407-12-138.
- Alexander, R. P., Fang, G., Rozowsky, J., Snyder, M. & Gerstein, M. B. 2010. Annotating non-coding regions of the genome. *Nature reviews Genetics* 11: 559-571.
- Alexandrov, L. B., Nik-Zainal, S., Wedge, D. C., Aparicio, S. A., Behjati, S., Biankin, A. V., Bignell, G. R., Bolli, N., Borg, A., Borresen-Dale, A. L., Boyault, S., Burkhardt, B., Butler, A. P., Caldas, C., Davies, H. R., Desmedt, C., Eils, R., Eyfjord, J. E., Foekens, J. A., Greaves, M., Hosoda, F., Hutter, B., Ilcic, T., Imbeaud, S., Imielinski, M., Jager, N., Jones, D. T., Jones, D., Knappskog, S., Kool, M., Lakhani, S. R., Lopez-Otin, C., Martin, S., Munshi, N. C., Nakamura, H., Northcott, P. A., Pajic, M., Papaemmanuil, E., Paradiso, A., Pearson, J. V., Puente, X. S., Raine, K., Ramakrishna, M., Richardson, A. L., Richter, J., Rosenstiel, P., Schlesner, M., Schumacher, T. N., Span, P. N., Teague, J. W., Totoki, Y., Tutt, A. N., Valdes-Mas, R., van Buuren, M. M., van 't Veer, L., Vincent-Salomon, A., Waddell, N., Yates, L. R., Australian Pancreatic Cancer Genome Initiative, ICGC Breast Cancer Consortium, ICGC MMML-Seq Consortium, ICGC PedBrain, Zucman-Rossi, J., Futreal, P. A., McDermott, U., Lichter, P., Meyerson, M., Grimmond, S. M., Siebert, R., Campo, E., Shibata, T., Pfister, S. M., Campbell, P. J. & Stratton, M. R. 2013. Signatures of mutational processes in human cancer. *Nature* 500: 415-421.
- Al-Mansour, M., Connors, J. M., Gascoyne, R. D., Skinnider, B. & Savage, K. J. 2010. Transformation to aggressive lymphoma in nodular lymphocyte-predominant Hodgkin's lymphoma. *Journal of clinical oncology: official journal of the American Society of Clinical Oncology* 28: 793-799.
- Almohideb, M., Watters, A. K. & Gerstein, W. 2013. Familial classic Kaposi sarcoma in two siblings: case report and literature review. *Journal of cutaneous medicine and surgery* 17: 356-361.
- Amakye, D., Jagani, Z. & Dorsch, M. 2013. Unraveling the therapeutic potential of the Hedgehog pathway in cancer. *Nature medicine* 19: 1410-1422.
- Amundadottir, L. T., Thorvaldsson, S., Gudbjartsson, D. F., Sulem, P., Kristjansson, K., Arnason, S., Gulcher, J. R., Bjornsson, J., Kong, A., Thorsteinsdottir, U. & Stefansson, K. 2004. Cancer as a complex phenotype: pattern of cancer distribution within and beyond the nuclear family. *PLoS medicine* 1: e65.
- Anagnostopoulos, I., Hansmann, M. L., Franssila, K., Harris, M., Harris, N. L., Jaffe, E. S., Han, J., van Krieken, J. M., Poppema, S., Marafioti, T., Franklin, J., Sextro, M., Diehl, V. & Stein, H. 2000. European Task Force on Lymphoma project on lymphocyte predominance Hodgkin disease: histologic and immunohistologic analysis of submitted cases reveals 2 types of Hodgkin disease with a nodular growth pattern and abundant lymphocytes. *Blood* 96: 1889-1899.
- Antinheimo, J., Sankila, R., Carpen, O., Pukkala, E., Sainio, M. & Jääskeläinen, J. 2000. Population-based analysis of sporadic and type 2 neurofibromatosis-associated meningiomas and schwannomas. *Neurology* 54: 71-76.
- Armitage, P. & Doll, R. 2004. The age distribution of cancer and a multi-stage theory of carcinogenesis. 1954. *International journal of epidemiology* 33: 1174-1179.
- Atayar, C., Kok, K., Kluiver, J., Bosga, A., van den Berg, E., van der Vlies, P., Blokzijl, T., Harms, G., Davelaar, I., Sikkema-Raddatz, B., Martin-Subero, J. I., Siebert, R., Poppema, S. & van den Berg, A. 2006. BCL6 alternative breakpoint region break and homozygous deletion of 17q24 in the nodular lymphocyte predominance type of Hodgkin's lymphoma-derived cell line DEV. *Human pathology* 37: 675-683.
- Avery, O. T., Macleod, C. M. & McCarty, M. 1944. Studies on the Chemical Nature of the Substance Inducing Transformation of Pneumococcal Types : Induction of Transformation by a Desoxyribonucleic Acid Fraction Isolated from Pneumococcus Type Iii. *The Journal of experimental medicine* 79: 137-158.

- Bacci, C., Sestini, R., Provenzano, A., Paganini, I., Mancini, I., Porfirio, B., Vivarelli, R., Genuardi, M. & Papi, L. 2010. Schwannomatosis associated with multiple meningiomas due to a familial SMARCB1 mutation. *Neurogenetics* 11: 73-80.
- Bai, M., Papoudou-Bai, A., Kitsoulis, P., Horianopoulos, N., Kamina, S., Agnantis, N. J. & Kanavaros, P. 2005. Cell cycle and apoptosis deregulation in classical Hodgkin lymphomas. *In vivo (Athens, Greece)* 19: 439-453.
- Bailey, J. A., Yavor, A. M., Massa, H. F., Trask, B. J. & Eichler, E. E. 2001. Segmental duplications: organization and impact within the current human genome project assembly. *Genome research* 11: 1005-1017.
- Bauduer, F., Vassallo, J., Delsol, G. & Brousset, P. 2005. Clustering and anticipation for nodular lymphocyte predominance Hodgkin lymphoma within a French Basque kindred. *British journal of haematology* 130: 648-649.
- Bauer, A. J. 2013. Clinical behavior and genetics of nonsyndromic, familial nonmedullary thyroid cancer. *Frontiers of hormone research* 41: 141-148.
- Beck, B. & Blanpain, C. 2013. Unravelling cancer stem cell potential. *Nature reviews cancer* 13: 727-738.
- Benjamini, Y. & Hochberg, Y. 1995. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B* 57(1): 289-300.
- Benson, G. 1999. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic acids research* 27: 573-580.
- Beral, V., Peterman, T. A., Berkelman, R. L. & Jaffe, H. W. 1990. Kaposi's sarcoma among persons with AIDS: a sexually transmitted infection? *The Lancet* 335: 123-128.
- Bianconi, E., Piovesan, A., Facchin, F., Beraudi, A., Casadei, R., Frabetti, F., Vitale, L., Pelleri, M. C., Tassani, S., Piva, F., Perez-Amodio, S., Strippoli, P. & Canaider, S. 2013. An estimation of the number of cells in the human body. *Annals of Human Biology* 40: 463-471.
- Biasoli, I., Stamatoullas, A., Meignin, V., Delmer, A., Reman, O., Morschhauser, F., Coiffier, B., Bosly, A., Divine, M. & Brice, P. 2010. Nodular, lymphocyte-predominant Hodgkin lymphoma: a long-term study and analysis of transformation to diffuse large B-cell lymphoma in a cohort of 164 patients from the Adult Lymphoma Study Group. *Cancer* 116: 631-639.
- Bolli, N., Avet-Loiseau, H., Wedge, D. C., Van Loo, P., Alexandrov, L. B., Martincorena, I., Dawson, K. J., Iorio, F., Nik-Zainal, S., Bignell, G. R., Hinton, J. W., Li, Y., Tubio, J. M., McLaren, S., O'Meara, S., Butler, A. P., Teague, J. W., Mudie, L., Anderson, E., Rashid, N., Tai, Y. T., Shammas, M. A., Sperling, A. S., Fulciniti, M., Richardson, P. G., Parmigiani, G., Magrangeas, F., Minvielle, S., Moreau, P., Attal, M., Facon, T., Futreal, P. A., Anderson, K. C., Campbell, P. J. & Munshi, N. C. 2014. Heterogeneity of genomic evolution and mutational profiles in multiple myeloma. *Nature communications* 5: 2997.
- Boshoff, C., Schulz, T. F., Kennedy, M. M., Graham, A. K., Fisher, C., Thomas, A., McGee, J. O., Weiss, R. A. & O'Leary, J. J. 1995. Kaposi's sarcoma-associated herpesvirus infects endothelial and spindle cells. *Nature medicine* 1: 1274-1278.
- Boström, J., Cobbers, J. M., Wolter, M., Tabatabai, G., Weber, R. G., Lichter, P., Collins, V. P. & Reifenberger, G. 1998. Mutation of the PTEN (MMAC1) tumor suppressor gene in a subset of glioblastomas but not in meningiomas with loss of chromosome arm 10q. *Cancer research* 58: 29-33.
- Bouvard, V., Baan, R., Straif, K., Grosse, Y., Secretan, B., El Ghissassi, F., Benbrahim-Tallaa, L., Guha, N., Freeman, C., Galichet, L., Coglian, V. & WHO International Agency for Research on Cancer Monograph Working Group. 2009. A review of human carcinogens--Part B: biological agents. *The Lancet oncology* 10: 321-322.
- Boveri, T. 1902. Über mehrpolige Mitosen als Mittel zur Analyse des Zellkerns. *Verh. D. Phys. Med. Ges. Wuzberg* N. F. 35:67-90.
- Boveri, T. 2008. Concerning the origin of malignant tumours by Theodor Boveri. Translated and annotated by Henry Harris. *Journal of cell science* 121 Suppl 1: 1-84.
- Boveri, T. 1914. Zur Frage der Entstehung maligner Tumoren. Jena: Fischer Verlag (English translation by Boveri M (1929). *The Origin of Malignant Tumors*. Baltimore: Waverly Press).
- Bozic, I., Antal, T., Ohtsuki, H., Carter, H., Kim, D., Chen, S., Karchin, R., Kinzler, K. W., Vogelstein, B. & Nowak, M. A. 2010. Accumulation of driver and passenger mutations during tumor progression. *Proceedings of the National Academy of Sciences of the United States of America* 107: 18545-18550.
- Brastianos, P. K., Horowitz, P. M., Santagata, S., Jones, R. T., McKenna, A., Getz, G., Ligon, K. L., Palescandolo, E., Van Hummelen, P., Ducar, M. D., Raza, A., Sunkavalli, A., Macconail, L. E., Stemmer-Rachamimov, A. O., Louis, D. N., Hahn, W. C., Dunn, I. F. & Beroukhi, R. 2013. Genomic sequencing of meningiomas identifies oncogenic SMO and AKT1 mutations. *Nature genetics* 45: 285-289.
- Broca, P.P. 1866. *Traité des tumeurs*. Paris, P. Asselin.
- Brouwer, J. R., Willemsen, R. & Oostra, B. A. 2009. Microsatellite repeat instability and neurological disease. *BioEssays : news and reviews in molecular, cellular and developmental biology* 31: 71-83.

- Brown, E. E., Fallin, D., Ruczinski, I., Hutchinson, A., Staats, B., Vitale, F., Lauria, C., Serraino, D., Rezza, G., Mbisa, G., Whitby, D., Messina, A., Goedert, J. J. & Chanock, S. J. 2006. Associations of classic Kaposi sarcoma with common variants in genes that modulate host immunity. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology* 15: 926-934.
- Brown, E. E., Fallin, M. D., Goedert, J. J., Chen, R., Whitby, D., Foster, C. B., Lauria, C., Alberg, A. J., Messina, A., Montella, M., Rezza, G., Vitale, F., Chanock, S. J. & Kaposi Sarcoma Genetics Working Group. 2005. A common genetic variant in FCGR3A-V158F and risk of Kaposi sarcoma herpesvirus infection and classic Kaposi sarcoma. *Cancer epidemiology, biomarkers & prevention: a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology* 14: 633-637.
- Byun, M., Abhyankar, A., Lelarge, V., Plancoulaine, S., Palanduz, A., Telhan, L., Boisson, B., Picard, C., Dewell, S., Zhao, C., Jouanguy, E., Feske, S., Abel, L. & Casanova, J. L. 2010. Whole-exome sequencing-based discovery of STIM1 deficiency in a child with fatal classic Kaposi sarcoma. *The Journal of experimental medicine* 207: 2307-2312.
- Byun, M., Ma, C. S., Akcay, A., Pedergnana, V., Palendira, U., Myoung, J., Avery, D. T., Liu, Y., Abhyankar, A., Lorenzo, L., Schmidt, M., Lim, H. K., Cassar, O., Migaud, M., Rozenberg, F., Canpolat, N., Aydogan, G., Fleckenstein, B., Bustamante, J., Picard, C., Gessain, A., Jouanguy, E., Cesarman, E., Olivier, M., Gros, P., Abel, L., Croft, M., Tangye, S. G. & Casanova, J. L. 2013. Inherited human OX40 deficiency underlying classic Kaposi sarcoma of childhood. *The Journal of experimental medicine* 210: 1743-1759.
- Callinan, P. A. & Batzer, M. A. 2006. Retrotransposable elements and human disease. *Genome dynamics* 1: 104-115.
- Camcioglu, Y., Picard, C., Lacoste, V., Dupuis, S., Akcakaya, N., Cokura, H., Kaner, G., Demirkesen, C., Plancoulaine, S., Emile, J. F., Gessain, A. & Casanova, J. L. 2004. HHV-8-associated Kaposi sarcoma in a child with IFN γ RI deficiency. *The Journal of pediatrics* 144: 519-523.
- Campbell, G. N., Lloyd, J., Wotherspoon, A., Coulter, C. & Bain, B. J. 2004. Nodular lymphocyte predominant Hodgkin lymphoma in siblings. *Leukemia & lymphoma* 45: 609-611.
- Cancer Genome Atlas Network. 2012. Comprehensive molecular characterization of human colon and rectal cancer. *Nature* 487: 330-337.
- Cannon-Albright, L. A., Thomas, A., Goldgar, D. E., Gholami, K., Rowe, K., Jacobsen, M., McWhorter, W. P. & Skolnick, M. H. 1994. Familiality of cancer in Utah. *Cancer research* 54: 2378-2385.
- Cartwright, R. A., Bernard, S. M., Bird, C. C., Darwin, C. M., O'Brien, C., Richards, I. D., Roberts, B. & McKinney, P. A. 1987. Chronic lymphocytic leukaemia: case control epidemiological study in Yorkshire. *British journal of cancer* 56: 79-82.
- Cazier, J. B. & Tomlinson, I. 2010. General lessons from large-scale studies to identify human cancer predisposition genes. *The Journal of pathology* 220: 255-262.
- Chang, Y., Cesarman, E., Pessin, M. S., Lee, F., Culpepper, J., Knowles, D. M. & Moore, P. S. 1994. Identification of herpesvirus-like DNA sequences in AIDS-associated Kaposi's sarcoma. *Science* 266: 1865-1869.
- Chapgier, A., Boisson-Dupuis, S., Jouanguy, E., Vogt, G., Feinberg, J., Prochnicka-Chalufour, A., Casrouge, A., Yang, K., Soudais, C., Fieschi, C., Santos, O. F., Bustamante, J., Picard, C., de Beaucoudrey, L., Emile, J. F., Arkwright, P. D., Schreiber, R. D., Rolinck-Werninghaus, C., Rosen-Wolff, A., Magdorf, K., Roesler, J. & Casanova, J. L. 2006. Novel STAT1 alleles in otherwise healthy patients with mycobacterial disease. *PLoS genetics* 2: e131.
- Chen, K., Wallis, J. W., McLellan, M. D., Larson, D. E., Kalicki, J. M., Pohl, C. S., McGrath, S. D., Wendl, M. C., Zhang, Q., Locke, D. P., Shi, X., Fulton, R. S., Ley, T. J., Wilson, R. K., Ding, L. & Mardis, E. R. 2009. BreakDancer: an algorithm for high-resolution mapping of genomic structural variation. *Nature methods* 6: 677-681.
- Cheng, S. Y. & Yue, S. 2008. Role and regulation of human tumor suppressor SUFU in Hedgehog signaling. *Advances in Cancer Research* 101: 29-43.
- Choi, J. D. & Lee, J. S. 2013. Interplay between Epigenetics and Genetics in Cancer. *Genomics & informatics* 11: 164-173.
- Christiaans, I., Kenter, S. B., Brink, H. C., van Os, T. A., Baas, F., van den Munckhof, P., Kidd, A. M. & Hulsebos, T. J. 2011. Germline SMARCB1 mutation and somatic NF2 mutations in familial multiple meningiomas. *Journal of medical genetics* 48: 93-97.
- Chung, C. C. & Chanock, S. J. 2011. Current status of genome-wide association studies in cancer. *Human genetics* 130: 59-78.
- Clark, V. E., Erson-Omay, E. Z., Serin, A., Yin, J., Cotney, J., Ozduman, K., Avsar, T., Li, J., Murray, P. B., Henegariu, O., Yilmaz, S., Gunel, J. M., Carrion-Grant, G., Yilmaz, B., Grady, C., Tanrikulu, B., Bakircioglu, M., Kaymakcalan, H., Caglayan, A. O., Sencar, L., Ceyhun, E., Atik, A. F., Bayri, Y., Bai, H., Kolb, L. E., Hebert, R. M., Omay, S. B., Mishra-Gorur, K., Choi, M., Overton, J. D., Holland, E. C., Mane, S., State, M. W.,

- Bilguvar, K., Baehring, J. M., Gutin, P. H., Piepmeyer, J. M., Vortmeyer, A., Brennan, C. W., Pamir, M. N., Kilic, T., Lifton, R. P., Noonan, J. P., Yasuno, K. & Gunel, M. 2013. Genomic analysis of non-NF2 meningiomas reveals mutations in TRAF7, KLF4, AKT1, and SMO. *Science* 339: 1077-1080.
- Colby, T. V., Hoppe, R. T. & Warnke, R. A. 1982. Hodgkin's disease: a clinicopathologic study of 659 cases. *Cancer* 49: 1848-1858.
- Coll-Mulet, L. & Gil, J. 2009. Genetic alterations in chronic lymphocytic leukaemia. *Clinical & translational oncology : official publication of the Federation of Spanish Oncology Societies and of the National Cancer Institute of Mexico* 11: 194-198.
- Comino-Mendez, I., Gracia-Aznarez, F. J., Schiavi, F., Landa, I., Leandro-Garcia, L. J., Leton, R., Honrado, E., Ramos-Medina, R., Caronia, D., Pita, G., Gomez-Grana, A., de Cubas, A. A., Inglada-Perez, L., Maliszewska, A., Taschin, E., Bobisse, S., Pica, G., Loli, P., Hernandez-Lavado, R., Diaz, J. A., Gomez-Morales, M., Gonzalez-Neira, A., Roncador, G., Rodriguez-Antona, C., Benitez, J., Mannelli, M., Opocher, G., Robledo, M. & Cascon, A. 2011. Exome sequencing identifies MAX mutations as a cause of hereditary pheochromocytoma. *Nature genetics* 43: 663-667.
- Cortez, D., Wang, Y., Qin, J. & Elledge, S. J. 1999. Requirement of ATM-dependent phosphorylation of brca1 in the DNA damage response to double-strand breaks. *Science* 286: 1162-1166.
- Crick, F. H., Barnett, L., Brenner, S. & Watts-Tobin, R. J. 1961. General nature of the genetic code for proteins. *Nature* 192: 1227-1232.
- Croce, C. M. 2008. Oncogenes and cancer. *The New England journal of medicine* 358: 502-511.
- Czene, K., Lichtenstein, P. & Hemminki, K. 2002. Environmental and heritable causes of cancer among 9.6 million individuals in the Swedish Family-Cancer Database. *International journal of cancer. Journal international du cancer* 99: 260-266.
- Dai, M., Wang, P., Boyd, A. D., Kostov, G., Athey, B., Jones, E. G., Bunney, W. E., Myers, R. M., Speed, T. P., Akil, H., Watson, S. J. & Meng, F. 2005. Evolving gene/transcript definitions significantly alter the interpretation of GeneChip data. *Nucleic acids research* 33: e175.
- Dean, M., Carrington, M., Winkler, C., Huttley, G. A., Smith, M. W., Allikmets, R., Goedert, J. J., Buchbinder, S. P., Vittinghoff, E., Gomperts, E., Donfield, S., Vlahov, D., Kaslow, R., Saah, A., Rinaldo, C., Detels, R. & O'Brien, S. J. 1996. Genetic restriction of HIV-1 infection and progression to AIDS by a deletion allele of the CKR5 structural gene. Hemophilia Growth and Development Study, Multicenter AIDS Cohort Study, Multicenter Hemophilia Cohort Study, San Francisco City Cohort, ALIVE Study. *Science (New York, N.Y.)* 273: 1856-1862.
- de Jong, D., Bosq, J., MacLennan, K. A., Diebold, J., Audouin, J., Chasle, J., Mandard, A. M., Marnay, J., Henry-Amar, M., European Organization for Research and Treatment of Cancer Lymphoma Group & Groupe d'Etude des Lymphomes de l'Adulte. 2006. Lymphocyte-rich classical Hodgkin lymphoma (LRCHL): clinico-pathological characteristics and outcome of a rare entity. *Annals of Oncology: Official Journal of the European Society for Medical Oncology / ESMO* 17: 141-145.
- de Martel, C., Ferlay, J., Franceschi, S., Vignat, J., Bray, F., Forman, D. & Plummer, M. 2012. Global burden of cancers attributable to infections in 2008: a review and synthetic analysis. *The Lancet oncology* 13: 607-615.
- Deng, H., Liu, R., Ellmeier, W., Choe, S., Unutmaz, D., Burkhart, M., Di Marzio, P., Marmon, S., Sutton, R. E., Hill, C. M., Davis, C. B., Peiper, S. C., Schall, T. J., Littman, D. R. & Landau, N. R. 1996. Identification of a major co-receptor for primary isolates of HIV-1. *Nature* 381: 661-666.
- DeLano, W. L. 2002. The PyMOL User's Manual.
- Dick, J. E. 2008. Stem cell concepts renew cancer research. *Blood* 112: 4793-4807.
- Dong, C. & Hemminki, K. 2001. Modification of cancer risks in offspring by sibling and parental cancers from 2,112,616 nuclear families. *International journal of cancer. Journal international du cancer* 92: 144-150.
- Doolittle, W. F. 2013. Is junk DNA bunk? A critique of ENCODE. *Proceedings of the National Academy of Sciences of the United States of America* 110: 5294-5300.
- Dorak, M. T., Yee, L. J., Tang, J., Shao, W., Lobashevsky, E. S., Jacobson, L. P. & Kaslow, R. A. 2005. HLA-B, -DRB1/3/4/5, and -DQB1 gene polymorphisms in human immunodeficiency virus-related Kaposi's sarcoma. *Journal of medical virology* 76: 302-310.
- Dragic, T., Litwin, V., Allaway, G. P., Martin, S. R., Huang, Y., Nagashima, K. A., Cayanan, C., Maddon, P. J., Koup, R. A., Moore, J. P. & Paxton, W. A. 1996. HIV-1 entry into CD4+ cells is mediated by the chemokine receptor CC-CKR-5. *Nature* 381: 667-673.
- Dupuy, A., Schulz, T., Chevret, S., Agbalika, F., Pellet, C., Janier, M., Dupin, N., Verola, O., Calvo, F. & Lebbe, C. 2009. Asymmetrical transmission of human herpesvirus 8 among spouses of patients with Kaposi sarcoma. *The British journal of dermatology* 160: 540-545.
- Easton, D. F. & Eeles, R. A. 2008. Genome-wide association studies in cancer. *Human molecular genetics* 17: R109-15.
- Eddy, S. R. 2012. The C-value paradox, junk DNA and ENCODE. *Current biology : CB* 22: R898-9.

- Edwards, S. L., Beesley, J., French, J. D. & Dunning, A. M. 2013. Beyond GWASs: illuminating the dark road from association to function. *American Journal of Human Genetics* **93**: 779-797.
- Emsley, P. & Cowtan, K. 2004. Coot: model-building tools for molecular graphics. *Acta crystallographica. Section D, Biological crystallography* **60**: 2126-2132.
- ENCODE Project Consortium, Bernstein, B. E., Birney, E., Dunham, I., Green, E. D., Gunter, C. & Snyder, M. 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**: 57-74.
- Eva, M. M., Yuki, K. E., Dauphinee, S. M., Schwartzentruber, J. A., Pyzik, M., Paquet, M., Lathrop, M., Majewski, J., Vidal, S. M. & Malo, D. 2014. Altered IFN-gamma-mediated immunity and transcriptional expression patterns in N-Ethyl-N-nitrosourea-induced STAT4 mutants confer susceptibility to acute typhoid-like disease. *Journal of immunology (Baltimore, Md.: 1950)* **192**: 259-270.
- Fearon, E. R. & Vogelstein, B. 1990. A genetic model for colorectal tumorigenesis. *Cell* **61**: 759-767.
- Ferlay, J., Soerjomataram, I., Ervik, M., Dikshit, R., Eser, S., Mathers, C., Rebelo, M., Parkin, D. M., Forman, D. & Bray, F. 2013. GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11 [Internet]. Available from: <http://globocan.iarc.fr>, accessed on day/month/year.
- Feuk, L., Carson, A. R. & Scherer, S. W. 2006. Structural variation in the human genome. *Nature reviews. Genetics* **7**: 85-97.
- Filipe-Santos, O., Bustamante, J., Chapgier, A., Vogt, G., de Beaucoudrey, L., Feinberg, J., Jouanguy, E., Boisson-Dupuis, S., Fieschi, C., Picard, C. & Casanova, J. L. 2006. Inborn errors of IL-12/23- and IFN-gamma-mediated immunity: molecular, cellular, and clinical features. *Seminars in immunology* **18**: 347-361.
- Fletcher, O. & Houlston, R. S. 2010. Architecture of inherited susceptibility to common cancer. *Nature reviews. Cancer* **10**: 353-361.
- Flint-Richter, P. & Sadetzki, S. 2007. Genetic predisposition for the development of radiation-associated meningioma: an epidemiological study. *The Lancet oncology* **8**: 403-410.
- Ford, A. M., Ridge, S. A., Cabrera, M. E., Mahmoud, H., Steel, C. M., Chan, L. C. & Greaves, M. 1993. In utero rearrangements in the trithorax-related oncogene in infant leukaemias. *Nature* **363**: 358-360.
- Forment, J. V., Kaidi, A. & Jackson, S. P. 2012. Chromothripsis and cancer: causes and consequences of chromosome shattering. *Nature reviews. Cancer* **12**: 663-670.
- Francis, J. M., Kiezun, A., Ramos, A. H., Serra, S., Pedamallu, C. S., Qian, Z. R., Banck, M. S., Kanwar, R., Kulkarni, A. A., Karpathakis, A., Manzo, V., Contractor, T., Philips, J., Nickerson, E., Pho, N., Hooshmand, S. M., Brais, L. K., Lawrence, M. S., Pugh, T., McKenna, A., Sivachenko, A., Cibulskis, K., Carter, S. L., Ojesina, A. I., Freeman, S., Jones, R. T., Voet, D., Saksena, G., Auclair, D., Onofrio, R., Shefler, E., Sougnez, C., Grimsby, J., Green, L., Lennon, N., Meyer, T., Caplin, M., Chung, D. C., Beutler, A. S., Ogino, S., Thirlwell, C., Shivdasani, R., Asa, S. L., Harris, C. R., Getz, G., Kulke, M. & Meyerson, M. 2013. Somatic mutation of CDKN1B in small intestine neuroendocrine tumors. *Nature genetics* **45**: 1483-1486.
- Friedman-Kien, A. E. 1981. Disseminated Kaposi's sarcoma syndrome in young homosexual men. *Journal of the American Academy of Dermatology* **5**: 468-471.
- Fung, Y. K., Murphree, A. L., T'Ang, A., Qian, J., Hinrichs, S. H. & Benedict, W. F. 1987. Structural evidence for the authenticity of the human retinoblastoma gene. *Science* **236**: 1657-1661.
- Gao, G., Bracken, A. P., Burkard, K., Pasini, D., Classon, M., Attwooll, C., Sagara, M., Imai, T., Helin, K. & Zhao, J. 2003. NPAT expression is regulated by E2F and is essential for cell cycle progression. *Molecular and cellular biology* **23**: 2821-2833.
- Genuardi, M. 2004. Inherited cancer predisposition. *American journal of medical genetics. Part C, Seminars in medical genetics* **129C**: 1-4.
- Gholami, K. & Thomas, A. 1994. A linear time algorithm for calculation of multiple pairwise kinship coefficients and the genetic index of familiarity. *Computers and biomedical research, an international journal* **27**: 342-350.
- Giles, G. G., Lickiss, J. N., Baikie, M. J., Lowenthal, R. M. & Panton, J. 1984. Myeloproliferative and lymphoproliferative disorders in Tasmania, 1972-80: occupational and familial aspects. *Journal of the National Cancer Institute* **72**: 1233-1240.
- Goldgar, D. E., Easton, D. F., Cannon-Albright, L. A. & Skolnick, M. H. 1994. Systematic population-based assessment of cancer risk in first-degree relatives of cancer probands. *Journal of the National Cancer Institute* **86**: 1600-1608.
- Goldin, L. R., Slager, S. L. & Caporaso, N. E. 2010. Familial chronic lymphocytic leukemia. *Current opinion in hematology* **17**: 350-355.
- Good, S. R., Thieu, V. T., Mathur, A. N., Yu, Q., Stritesky, G. L., Yeh, N., O'Malley, J. T., Perumal, N. B. & Kaplan, M. H. 2009. Temporal induction pattern of STAT4 target genes defines potential for Th1 lineage-specific programming. *Journal of immunology (Baltimore, Md.: 1950)* **183**: 3839-3847.
- Gorlin, R. J. 1987. Nevroid basal-cell carcinoma syndrome. *Medicine* **66**: 98-113.

- Graur, D., Zheng, Y., Price, N., Azevedo, R. B., Zufall, R. A. & Elhaik, E. 2013. On the immortality of television sets: "function" in the human genome according to the evolution-free gospel of ENCODE. *Genome biology and evolution* 5: 578-590.
- Gryfe, R. & Gallinger, S. 2001. Microsatellite instability, mismatch repair deficiency, and colorectal cancer. *Surgery* 130: 17-20.
- Gunz, F. W., Gunz, J. P., Veale, A. M., Chapman, C. J. & Houston, I. B. 1975. Familial leukaemia: a study of 909 families. *Scandinavian Journal of Haematology* 15: 117-131.
- Guttman-Yassky, E., Cohen, A., Kra-Oz, Z., Friedman-Birnbaum, R., Sprecher, E., Zaltzman, N., Friedman, E., Silbermann, M., Rubin, D., Linn, S., Whitby, D., Gideoni, O., Pollack, S., Bergman, R. & Sarid, R. 2004. Familial clustering of classic Kaposi sarcoma. *The Journal of infectious diseases* 189: 2023-2026.
- Guttman-Yassky, E., Kra-Oz, Z., Dubnov, J., Friedman-Birnbaum, R., Segal, I., Zaltzman, N., Roth, T., Schwartz, F., Linn, S., Rozenman, D., David, M., Silbermann, M., Barchana, M., Bergman, R. & Sarid, R. 2005. Infection with Kaposi's sarcoma-associated herpesvirus among families of patients with classic Kaposi's sarcoma. *Archives of Dermatology* 141: 1429-1434.
- Hanahan, D. & Weinberg, R. A. 2011. Hallmarks of cancer: the next generation. *Cell* 144: 646-674.
- Hanahan, D. & Weinberg, R. A. 2000. The hallmarks of cancer. *Cell* 100: 57-70.
- Hanks, S., Coleman, K., Reid, S., Plaja, A., Firth, H., Fitzpatrick, D., Kidd, A., Mehes, K., Nash, R., Robin, N., Shannon, N., Tolmie, J., Swansbury, J., Irrthum, A., Douglas, J. & Rahman, N. 2004. Constitutional aneuploidy and cancer predisposition caused by biallelic mutations in BUB1B. *Nature genetics* 36: 1159-1161.
- Hao, H. X., Khalimonchuk, O., Schraders, M., Dephoure, N., Bayley, J. P., Kunst, H., Devilee, P., Cremers, C. W., Schiffman, J. D., Bentz, B. G., Gygi, S. P., Winge, D. R., Kremer, H. & Rutter, J. 2009. SDH5, a gene required for flavination of succinate dehydrogenase, is mutated in paraganglioma. *Science* 325: 1139-1142.
- Harmston, N. & Lenhard, B. 2013. Chromatin and epigenetic features of long-range gene regulation. *Nucleic acids research* 41: 7185-7199.
- Harris, N. L., Jaffe, E. S., Stein, H., Banks, P. M., Chan, J. K., Cleary, M. L., Delsol, G., De Wolf-Peeters, C., Falini, B. & Gatter, K. C. 1994. A revised European-American classification of lymphoid neoplasms: a proposal from the International Lymphoma Study Group. *Blood* 84: 1361-1392.
- Heinrich, B., Hartmann, C., Stemmer-Rachamimov, A. O., Louis, D. N. & MacCollin, M. 2003. Multiple meningiomas: Investigating the molecular basis of sporadic and familial forms. *International journal of cancer* 103: 483-488.
- Hemminki, K., Dong, C. & Vaitinen, P. 2001. Cancer risks to spouses and offspring in the Family-Cancer Database. *Genetic epidemiology* 20: 247-257.
- Hemminki, K. & Jiang, Y. 2002. Cancer risks among long-standing spouses. *British journal of cancer* 86: 1737-1740.
- Hemminki, K. & Li, X. 2001. Familial carcinoid tumors and subsequent cancers: a nation-wide epidemiologic study from Sweden. *International journal of cancer* 94: 444-448.
- Hemminki, K., Lorenzo Bermejo, J. & Forsti, A. 2006. The balance between heritable and environmental aetiology of human disease. *Nature reviews. Genetics* 7: 958-965.
- Hindorff, L. A., Sethupathy, P., Junkins, H. A., Ramos, E. M., Mehta, J. P., Collins, F. S. & Manolio, T. A. 2009. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proceedings of the National Academy of Sciences of the United States of America* 106: 9362-9367.
- Hiripi, E., Bermejo, J. L., Sundquist, J. & Hemminki, K. 2009. Familial gastrointestinal carcinoid tumours and associated cancers. *Annals of Oncology : Official Journal of the European Society for Medical Oncology / ESMO* 20: 950-954.
- Hirschhorn, J. N. & Daly, M. J. 2005. Genome-wide association studies for common diseases and complex traits. *Nature reviews. Genetics* 6: 95-108.
- Hjalgrim, H., Melbye, M., Pukkala, E., Langmark, F., Frisch, M., Dictor, M. & Ekblom, A. 1996. Epidemiology of Kaposi's sarcoma in the Nordic countries before the AIDS epidemic. *British journal of cancer* 74: 1499-1502.
- Hollestelle, A., Wasielewski, M., Martens, J. W. & Schutte, M. 2010. Discovering moderate-risk breast cancer susceptibility genes. *Current opinion in genetics & development* 20: 268-276.
- Hoogstraat, M., de Pagter, M. S., Cirkel, G. A., van Roosmalen, M. J., Harkins, T. T., Duran, K., Kreeftmeijer, J., Renkens, I., Witteveen, P. O., Lee, C. C., Nijman, I. J., Guy, T., van 't Slot, R., Jonges, T. N., Lolkema, M. P., Koudijs, M. J., Zweemer, R. P., Voest, E. E., Cuppen, E. & Kloosterman, W. P. 2014. Genomic and transcriptomic plasticity in treatment-naïve ovarian cancer. *Genome research* 24: 200-211.
- Huang, H., Buhl, R., Hugo, H. H. & Mehdorn, H. M. 2005. Clinical and histological features of multiple meningiomas compared with solitary meningiomas. *Neurological research* 27: 324-332.
- Hussain, S. K., Sundquist, J. & Hemminki, K. 2008. Familial clustering of cancer at human papillomavirus-associated sites according to the Swedish Family-Cancer Database. *International journal of cancer. Journal international du cancer* 122: 1873-1878.

- IARC Working group. 2012a. IARC monographs on the evaluation of the carcinogenic risk to humans, biological agents, review of human carcinogen. 100B: 498.
- IARC Working group. 2012b. IARC monographs on the evaluation of the carcinogenic risk to humans, biological agents, review of human carcinogen: Kaposi Sarcoma Herpesvirus. 100B: 498-169-196.
- Imai, T., Yamauchi, M., Seki, N., Sugawara, T., Saito, T., Matsuda, Y., Ito, H., Nagase, T., Nomura, N. & Hori, T. 1996. Identification and characterization of a new gene physically linked to the ATM gene. *Genome research* 6: 439-447.
- Ingham, P. W., Nakano, Y. & Seger, C. 2011. Mechanisms and functions of Hedgehog signalling across the metazoa. *Nature reviews.Genetics* 12: 393-406.
- International Cancer Genome Consortium, Hudson, T. J., Anderson, W., Artez, A., Barker, A. D., Bell, C., Bernabe, R. R., Bhan, M. K., Calvo, F., Eerola, I., Gerhard, D. S., Guttmacher, A., Guyer, M., Hemsley, F. M., Jennings, J. L., Kerr, D., Klatt, P., Kolar, P., Kusada, J., Lane, D. P., Laplace, F., Youyong, L., Nettekoven, G., Ozenberger, B., Peterson, J., Rao, T. S., Remacle, J., Schafer, A. J., Shibata, T., Stratton, M. R., Vockley, J. G., Watanabe, K., Yang, H., Yuen, M. M., Knoppers, B. M., Bobrow, M., Cambon-Thomsen, A., Dressler, L. G., Dyke, S. O., Joly, Y., Kato, K., Kennedy, K. L., Nicolas, P., Parker, M. J., Rial-Sebbag, E., Romeo-Casabona, C. M., Shaw, K. M., Wallace, S., Wiesner, G. L., Zeps, N., Lichter, P., Biankin, A. V., Chabannon, C., Chin, L., Clement, B., de Alava, E., Degos, F., Ferguson, M. L., Geary, P., Hayes, D. N., Hudson, T. J., Johns, A. L., Kasprzyk, A., Nakagawa, H., Penny, R., Piris, M. A., Sarin, R., Scarpa, A., Shibata, T., van de Vijver, M., Futreal, P. A., Aburatani, H., Bayes, M., Botwell, D. D., Campbell, P. J., Estivill, X., Gerhard, D. S., Grimmond, S. M., Gut, I., Hirst, M., Lopez-Otin, C., Majumder, P., Marra, M., McPherson, J. D., Nakagawa, H., Ning, Z., Puente, X. S., Ruan, Y., Shibata, T., Stratton, M. R., Stunnenberg, H. G., Swerdlow, H., Velculescu, V. E., Wilson, R. K., Xue, H. H., Yang, L., Spellman, P. T., Bader, G. D., Boutros, P. C., Campbell, P. J., Flicek, P., Getz, G., Guigo, R., Guo, G., Haussler, D., Heath, S., Hubbard, T. J., Jiang, T., Jones, S. M., Li, Q., Lopez-Bigas, N., Luo, R., Muthuswamy, L., Ouellette, B. F., Pearson, J. V., Puente, X. S., Quesada, V., Raphael, B. J., Sander, C., Shibata, T., Speed, T. P., Stein, L. D., Stuart, J. M., Teague, J. W., Totoki, Y., Tsunoda, T., Valencia, A., Wheeler, D. A., Wu, H., Zhao, S., Zhou, G., Stein, L. D., Guigo, R., Hubbard, T. J., Joly, Y., Jones, S. M., Kasprzyk, A., Lathrop, M., Lopez-Bigas, N., Ouellette, B. F., Spellman, P. T., Teague, J. W., Thomas, G., Valencia, A., Yoshida, T., Kennedy, K. L., Axton, M., Dyke, S. O., Futreal, P. A., Gerhard, D. S., Gunter, C., Guyer, M., Hudson, T. J., McPherson, J. D., Miller, L. J., Ozenberger, B., Shaw, K. M., Kasprzyk, A., Stein, L. D., Zhang, J., Haider, S. A., Wang, J., Yung, C. K., Cros, A., Liang, Y., Gnaneshan, S., Guberman, J., Hsu, J., Bobrow, M., Chalmers, D. R., Hasel, K. W., Joly, Y., Kaan, T. S., Kennedy, K. L., Knoppers, B. M., Lowrance, W. W., Masui, T., Nicolas, P., Rial-Sebbag, E., Rodriguez, L. L., Vergely, C., Yoshida, T., Grimmond, S. M., Biankin, A. V., Bowtell, D. D., Cloonan, N., deFazio, A., Eshleman, J. R., Etemadmoghadam, D., Gardiner, B. B., Kench, J. G., Scarpa, A., Sutherland, R. L., Tempero, M. A., Waddell, N. J., Wilson, P. J., McPherson, J. D., Gallinger, S., Tsao, M. S., Shaw, P. A., Petersen, G. M., Mukhopadhyay, D., Chin, L., DePinho, R. A., Thayer, S., Muthuswamy, L., Shazand, K., Beck, T., Sam, M., Timms, L., Ballin, V., Lu, Y., Ji, J., Zhang, X., Chen, F., Hu, X., Zhou, G., Yang, Q., Tian, G., Zhang, L., Xing, X., Li, X., Zhu, Z., Yu, Y., Yu, J., Yang, H., Lathrop, M., Tost, J., Brennan, P., Holcatova, I., Zaridze, D., Brazma, A., Egevard, L., Prokhorchouk, E., Banks, R. E., Uhlen, M., Cambon-Thomsen, A., Viksna, J., Ponten, F., Skryabin, K., Stratton, M. R., Futreal, P. A., Birney, E., Borg, A., Borresen-Dale, A. L., Caldas, C., Foekens, J. A., Martin, S., Reis-Filho, J. S., Richardson, A. L., Sotiriou, C., Stunnenberg, H. G., Thoms, G., van de Vijver, M., van't Veer, L., Calvo, F., Birnbaum, D., Blanche, H., Boucher, P., Boyault, S., Chabannon, C., Gut, I., Masson-Jacquemier, J. D., Lathrop, M., Pauporte, I., Pivot, X., Vincent-Salomon, A., Tabone, E., Theillet, C., Thomas, G., Tost, J., Treilleux, I., Calvo, F., Bioulac-Sage, P., Clement, B., Decaens, T., Degos, F., Franco, D., Gut, I., Gut, M., Heath, S., Lathrop, M., Samuel, D., Thomas, G., Zucman-Rossi, J., Lichter, P., Eils, R., Brors, B., Korbel, J. O., Korshunov, A., Landgraf, P., Lehrach, H., Pfister, S., Radlwimmer, B., Reifemberger, G., Taylor, M. D., von Kalle, C., Majumder, P. P., Sarin, R., Rao, T. S., Bhan, M. K., Scarpa, A., Pederzoli, P., Lawlor, R. A., Delledonne, M., Bardelli, A., Biankin, A. V., Grimmond, S. M., Gress, T., Klimstra, D., Zamboni, G., Shibata, T., Nakamura, Y., Nakagawa, H., Kusada, J., Tsunoda, T., Miyano, S., Aburatani, H., Kato, K., Fujimoto, A., Yoshida, T., Campo, E., Lopez-Otin, C., Estivill, X., Guigo, R., de Sanjose, S., Piris, M. A., Montserrat, E., Gonzalez-Diaz, M., Puente, X. S., Jares, P., Valencia, A., Himmelbauer, H., Quesada, V., Bea, S., Stratton, M. R., Futreal, P. A., Campbell, P. J., Vincent-Salomon, A., Richardson, A. L., Reis-Filho, J. S., van de Vijver, M., Thomas, G., Masson-Jacquemier, J. D., Aparicio, S., Borg, A., Borresen-Dale, A. L., Caldas, C., Foekens, J. A., Stunnenberg, H. G., van't Veer, L., Easton, D. F., Spellman, P. T., Martin, S., Barker, A. D., Chin, L., Collins, F. S., Compton, C. C., Ferguson, M. L., Gerhard, D. S., Getz, G., Gunter, C., Guttmacher, A., Guyer, M., Hayes, D. N., Lander, E. S., Ozenberger, B., Penny, R., Peterson, J., Sander, C., Shaw, K. M., Speed, T. P., Spellman, P. T., Vockley, J. G., Wheeler, D. A., Wilson, R. K., Hudson, T. J., Chin, L., Knoppers, B. M., Lander, E. S., Lichter, P., Stein, L. D., Stratton, M. R., Anderson, W., Barker, A. D., Bell, C., Bobrow, M., Burke, W., Collins,

- F. S., Compton, C. C., DePinho, R. A., Easton, D. F., Futreal, P. A., Gerhard, D. S., Green, A. R., Guyer, M., Hamilton, S. R., Hubbard, T. J., Kallioniemi, O. P., Kennedy, K. L., Ley, T. J., Liu, E. T., Lu, Y., Majumder, P., Marra, M., Ozenberger, B., Peterson, J., Schafer, A. J., Spellman, P. T., Stunnenberg, H. G., Wainwright, B. J., Wilson, R. K. & Yang, H. 2010. International network of cancer genome projects. *Nature* 464: 993-998.
- International HapMap Consortium. 2003. The International HapMap Project. *Nature* 426: 789-796.
- International HapMap Consortium. 2005. A haplotype map of the human genome. *Nature* 437: 1299-1320.
- International HapMap Consortium, Frazer, K. A., Ballinger, D. G., Cox, D. R., Hinds, D. A., Stuve, L. L., Gibbs, R. A., Belmont, J. W., Boudreau, A., Hardenbol, P., Leal, S. M., Pasternak, S., Wheeler, D. A., Willis, T. D., Yu, F., Yang, H., Zeng, C., Gao, Y., Hu, H., Hu, W., Li, C., Lin, W., Liu, S., Pan, H., Tang, X., Wang, J., Wang, W., Yu, J., Zhang, B., Zhang, Q., Zhao, H., Zhao, H., Zhou, J., Gabriel, S. B., Barry, R., Blumenstiel, B., Camargo, A., Defelice, M., Faggart, M., Goyette, M., Gupta, S., Moore, J., Nguyen, H., Onofrio, R. C., Parkin, M., Roy, J., Stahl, E., Winchester, E., Ziaugra, L., Altshuler, D., Shen, Y., Yao, Z., Huang, W., Chu, X., He, Y., Jin, L., Liu, Y., Shen, Y., Sun, W., Wang, H., Wang, Y., Wang, Y., Xiong, X., Xu, L., Wayne, M. M., Tsui, S. K., Xue, H., Wong, J. T., Galver, L. M., Fan, J. B., Gunderson, K., Murray, S. S., Oliphant, A. R., Chee, M. S., Montpetit, A., Chagnon, F., Ferretti, V., Leboeuf, M., Olivier, J. F., Phillips, M. S., Roumy, S., Sallee, C., Verner, A., Hudson, T. J., Kwok, P. Y., Cai, D., Koboldt, D. C., Miller, R. D., Pawlikowska, L., Taillon-Miller, P., Xiao, M., Tsui, L. C., Mak, W., Song, Y. Q., Tam, P. K., Nakamura, Y., Kawaguchi, T., Kitamoto, T., Morizono, T., Nagashima, A., Ohnishi, Y., Sekine, A., Tanaka, T., Tsunoda, T., Deloukas, P., Bird, C. P., Delgado, M., Dermitzakis, E. T., Gwilliam, R., Hunt, S., Morrison, J., Powell, D., Stranger, B. E., Whittaker, P., Bentley, D. R., Daly, M. J., de Bakker, P. I., Barrett, J., Chretien, Y. R., Maller, J., McCarroll, S., Patterson, N., Pe'er, I., Price, A., Purcell, S., Richter, D. J., Sabeti, P., Saxena, R., Schaffner, S. F., Sham, P. C., Varilly, P., Altshuler, D., Stein, L. D., Krishnan, L., Smith, A. V., Tello-Ruiz, M. K., Thorisson, G. A., Chakravarti, A., Chen, P. E., Cutler, D. J., Kashuk, C. S., Lin, S., Abecasis, G. R., Guan, W., Li, Y., Munro, H. M., Qin, Z. S., Thomas, D. J., McVean, G., Auton, A., Bottolo, L., Cardin, N., Eyheramendy, S., Freeman, C., Marchini, J., Myers, S., Spencer, C., Stephens, M., Donnelly, P., Cardon, L. R., Clarke, G., Evans, D. M., Morris, A. P., Weir, B. S., Tsunoda, T., Mullikin, J. C., Sherry, S. T., Feolo, M., Skol, A., Zhang, H., Zeng, C., Zhao, H., Matsuda, I., Fukushima, Y., Macer, D. R., Suda, E., Rotimi, C. N., Adebamowo, C. A., Ajayi, I., Aniagwu, T., Marshall, P. A., Nkwdimmah, C., Royal, C. D., Leppert, M. F., Dixon, M., Peiffer, A., Qiu, R., Kent, A., Kato, K., Niikawa, N., Adewole, I. F., Knoppers, B. M., Foster, M. W., Clayton, E. W., Watkin, J., Gibbs, R. A., Belmont, J. W., Muzny, D., Nazareth, L., Sodergren, E., Weinstock, G. M., Wheeler, D. A., Yakub, I., Gabriel, S. B., Onofrio, R. C., Richter, D. J., Ziaugra, L., Birren, B. W., Daly, M. J., Altshuler, D., Wilson, R. K., Fulton, L. L., Rogers, J., Burton, J., Carter, N. P., Clee, C. M., Griffiths, M., Jones, M. C., McLay, K., Plumb, R. W., Ross, M. T., Sims, S. K., Willey, D. L., Chen, Z., Han, H., Kang, L., Godbout, M., Wallenburg, J. C., L'Archeveque, P., Bellemare, G., Saeki, K., Wang, H., An, D., Fu, H., Li, Q., Wang, Z., Wang, R., Holden, A. L., Brooks, L. D., McEwen, J. E., Guyer, M. S., Wang, V. O., Peterson, J. L., Shi, M., Spiegel, J., Sung, L. M., Zacharia, L. F., Collins, F. S., Kennedy, K., Jamieson, R. & Stewart, J. 2007. A second generation human haplotype map of over 3.1 million SNPs. *Nature* 449: 851-861.
- International Human Genome Sequencing Consortium. 2004. Finishing the euchromatic sequence of the human genome. *Nature* 431: 931-945.
- INTERPHONE Study Group. 2010. Brain tumour risk in relation to mobile telephone use: results of the INTERPHONE international case-control study. *International journal of epidemiology* 39: 675-694.
- Iwasaki, M., Mukai, T., Nakajima, C., Yang, Y. F., Gao, P., Yamaguchi, N., Tomura, M., Fujiwara, H. & Hamaoka, T. 2001. A mandatory role for STAT4 in IL-12 induction of mouse T cell CCR5. *Journal of immunology (Baltimore, Md.: 1950)* 167: 6877-6883.
- Jakkula, E., Rehnstrom, K., Varilo, T., Pietilainen, O. P., Paunio, T., Pedersen, N. L., deFaire, U., Jarvelin, M. R., Saharinen, J., Freimer, N., Ripatti, S., Purcell, S., Collins, A., Daly, M. J., Palotie, A. & Peltonen, L. 2008. The genome-wide patterns of variation expose significant substructure in a founder population. *American Journal of Human Genetics* 83: 787-794.
- Jeghers, H., Mc Kusick, V. A. & Katz, K. H. 1949. Generalized intestinal polyposis and melanin spots of the oral mucosa, lips and digits; a syndrome of diagnostic significance. *The New England journal of medicine* 241: 993, illust; passim.
- Jiang, D. K., Sun, J., Cao, G., Liu, Y., Lin, D., Gao, Y. Z., Ren, W. H., Long, X. D., Zhang, H., Ma, X. P., Wang, Z., Jiang, W., Chen, T. Y., Gao, Y., Sun, L. D., Long, J. R., Huang, H. X., Wang, D., Yu, H., Zhang, P., Tang, L. S., Peng, B., Cai, H., Liu, T. T., Zhou, P., Liu, F., Lin, X., Tao, S., Wan, B., Sai-Yin, H. X., Qin, L. X., Yin, J., Liu, L., Wu, C., Pei, Y., Zhou, Y. F., Zhai, Y., Lu, P. X., Tan, A., Zuo, X. B., Fan, J., Chang, J., Gu, X., Wang, N. J., Li, Y., Liu, Y. K., Zhai, K., Zhang, H., Hu, Z., Liu, J., Yi, Q., Xiang, Y., Shi, R., Ding, Q., Zheng, W., Shu, X. O., Mo, Z., Shugart, Y. Y., Zhang, X. J., Zhou, G., Shen, H., Zheng, S. L., Xu, J. & Yu, L. 2013.

Genetic variants in STAT4 and HLA-DQ genes confer risk of hepatitis B virus-related hepatocellular carcinoma. *Nature genetics* 45: 72-75.

- Jones, D. T., Jager, N., Kool, M., Zichner, T., Hutter, B., Sultan, M., Cho, Y. J., Pugh, T. J., Hovestadt, V., Stutz, A. M., Rausch, T., Warnatz, H. J., Ryzhova, M., Bender, S., Sturm, D., Pleier, S., Cin, H., Pfaff, E., Sieber, L., Wittmann, A., Remke, M., Witt, H., Hutter, S., Tzaridis, T., Weischenfeldt, J., Raeder, B., Avci, M., Amstislavskiy, V., Zapatka, M., Weber, U. D., Wang, Q., Lasitschka, B., Bartholomae, C. C., Schmidt, M., von Kalle, C., Ast, V., Lawerenz, C., Eils, J., Kabbe, R., Benes, V., van Sluis, P., Koster, J., Volckmann, R., Shih, D., Betts, M. J., Russell, R. B., Coco, S., Tonini, G. P., Schuller, U., Hans, V., Graf, N., Kim, Y. J., Monoranu, C., Roggendorf, W., Unterberg, A., Herold-Mende, C., Milde, T., Kulozik, A. E., von Deimling, A., Witt, O., Maass, E., Rossler, J., Ebinger, M., Schuhmann, M. U., Fruhwald, M. C., Hasselblatt, M., Jabado, N., Rutkowski, S., von Bueren, A. O., Williamson, D., Clifford, S. C., McCabe, M. G., Collins, V. P., Wolf, S., Wiemann, S., Lehrach, H., Brors, B., Scheurlen, W., Felsberg, J., Reifemberger, G., Northcott, P. A., Taylor, M. D., Meyerson, M., Pomeroy, S. L., Yaspo, M. L., Korbel, J. O., Korshunov, A., Eils, R., Pfister, S. M. & Lichter, P. 2012. Dissecting the genomic complexity underlying medulloblastoma. *Nature* 488: 100-105.
- Kaasinen, E., Aavikko, M., Vahteristo, P., Patama, T., Li, Y., Saarinen, S., Kilpivaara, O., Pitkanen, E., Knekt, P., Laaksonen, M., Artama, M., Lehtonen, R., Aaltonen, L. A. & Pukkala, E. 2013. Nationwide registry-based analysis of cancer clustering detects strong familial occurrence of Kaposi sarcoma. *PloS one* 8: e55209.
- Kalla, C., Scheuermann, M. O., Kube, I., Schlotter, M., Mertens, D., Dohner, H., Stilgenbauer, S. & Lichter, P. 2007. Analysis of 11q22-q23 deletion target genes in B-cell chronic lymphocytic leukaemia: evidence for a pathogenic role of NPAT, CUL5, and PPP2R1B. *European journal of cancer (Oxford, England : 1990)* 43: 1328-1335.
- Kaplan, M. H. 2005. STAT4: a critical regulator of inflammation in vivo. *Immunologic research* 31: 231-242.
- Kaplan, M. H., Sun, Y. L., Hoey, T. & Grusby, M. J. 1996. Impaired IL-12 responses and enhanced development of Th2 cells in Stat4-deficient mice. *Nature* 382: 174-177.
- Kaposi, M. 1872. Idiopathisches multiples Pigmentsarkom der Haut. 4:
- Kauppi, L., Jeffreys, A. J. & Keeney, S. 2004. Where the crossovers are: recombination distributions in mammals. *Nature reviews. Genetics* 5: 413-424.
- Kaur, G., Sharma, G., Kumar, N., Kaul, M. H., Bansal, R. A., Vajpayee, M., Wig, N., Sharma, S. K. & Mehra, N. K. 2013. Genomic architecture of HIV-1 infection: current status & challenges. *The Indian journal of medical research* 138: 663-681.
- Kharazmi, E., Pukkala, E., Sundquist, K. & Hemminki, K. 2013. Familial risk of small intestinal carcinoid and adenocarcinoma. *Clinical gastroenterology and hepatology : the official clinical practice journal of the American Gastroenterological Association* 11: 944-949.
- Kidd, J. M., Cooper, G. M., Donahue, W. F., Hayden, H. S., Sampas, N., Graves, T., Hansen, N., Teague, B., Alkan, C., Antonacci, F., Haugen, E., Zerr, T., Yamada, N. A., Tsang, P., Newman, T. L., Tuzun, E., Cheng, Z., Ebling, H. M., Tusneem, N., David, R., Gillett, W., Phelps, K. A., Weaver, M., Saranga, D., Brand, A., Tao, W., Gustafson, E., McKernan, K., Chen, L., Malig, M., Smith, J. D., Korn, J. M., McCarroll, S. A., Altshuler, D. A., Peiffer, D. A., Dorschner, M., Stamatoyannopoulos, J., Schwartz, D., Nickerson, D. A., Mullikin, J. C., Wilson, R. K., Bruhn, L., Olson, M. V., Kaul, R., Smith, D. R. & Eichler, E. E. 2008. Mapping and sequencing of structural variation from eight human genomes. *Nature* 453: 56-64.
- Kijima, C., Miyashita, T., Suzuki, M., Oka, H. & Fujii, K. 2012. Two cases of nevoid basal cell carcinoma syndrome associated with meningioma caused by a PTCH1 or SUFU germline mutation. *Familial cancer* 11: 565-570.
- Kinzler, K. W. & Vogelstein, B. 1998. Landscaping the cancer terrain. *Science* 280: 1036-1037.
- Kinzler, K. W. & Vogelstein, B. 1997. Cancer-susceptibility genes. Gatekeepers and caretakers. *Nature* 386: 761, 763.
- Klaeboe, L., Lonn, S., Scheie, D., Auvinen, A., Christensen, H. C., Feychting, M., Johansen, C., Salminen, T. & Tynes, T. 2005. Incidence of intracranial meningiomas in Denmark, Finland, Norway and Sweden, 1968-1997. *International journal of cancer. Journal international du cancer* 117: 996-1001.
- Knudson, A. G., Jr. 1971. Mutation and cancer: statistical study of retinoblastoma. *Proceedings of the National Academy of Sciences of the United States of America* 68: 820-823.
- Kogerman, P., Grimm, T., Kogerman, L., Krause, D., Unden, A. B., Sandstedt, B., Toftgard, R. & Zaphiropoulos, P. G. 1999. Mammalian suppressor-of-fused modulates nuclear-cytoplasmic shuttling of Gli-1. *Nature cell biology* 1: 312-319.
- Koressaar, T. & Remm, M. 2007. Enhancements and modifications of primer design program Primer3. *Bioinformatics (Oxford, England)* 23: 1289-1291.
- Krampla, W., Newrkla, S., Pfisterer, W., Jungwirth, S., Fischer, P., Leitha, T., Hruby, W. & Tragl, K. H. 2004. Frequency and risk factors for meningioma in clinically healthy 75-year-old patients: results of the Transdanube Ageing Study (VITA). *Cancer* 100: 1208-1212.

- Kruglyak, L., Daly, M. J., Reeve-Daly, M. P. & Lander, E. S. 1996. Parametric and nonparametric linkage analysis: a unified multipoint approach. *American Journal of Human Genetics* **58**: 1347-1363.
- Kumar, P., Henikoff, S. & Ng, P. C. 2009. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nature protocols* **4**: 1073-1081.
- Kuppers, R. 2009. The biology of Hodgkin's lymphoma. *Nature reviews.Cancer* **9**: 15-27.
- Lada, A. G., Dhar, A., Boissy, R. J., Hirano, M., Rubel, A. A., Rogozin, I. B. & Pavlov, Y. I. 2012. AID/APOBEC cytosine deaminase induces genome-wide kataegis. *Biology direct* **7**: 47; discussion 47-6150-7-47.
- Lahiri, D. K. & Nurnberger, J. I., Jr. 1991. A rapid non-enzymatic method for the preparation of HMW DNA from blood for RFLP studies. *Nucleic acids research* **19**: 5444.
- Lai, K., Kaspar, B. K., Gage, F. H. & Schaffer, D. V. 2003. Sonic hedgehog regulates adult neural progenitor proliferation in vitro and in vivo. *Nature neuroscience* **6**: 21-27.
- Laken, S. J., Petersen, G. M., Gruber, S. B., Oddoux, C., Ostrer, H., Giardiello, F. M., Hamilton, S. R., Hampel, H., Markowitz, A., Klimstra, D., Jhanwar, S., Winawer, S., Offit, K., Luce, M. C., Kinzler, K. W. & Vogelstein, B. 1997. Familial colorectal cancer in Ashkenazim due to a hypermutable tract in APC. *Nature genetics* **17**: 79-83.
- Lall, R. R., Lall, R. R., Smith, T. R., Lee, K. H., Mao, Q., Kalapurakal, J. A., Marymont, M. H. & Chandler, J. P. 2013. Delayed malignant transformation of petroclival meningioma to chondrosarcoma after stereotactic radiosurgery. *Journal of clinical neuroscience : official journal of the Neurosurgical Society of Australasia*
- Laloo, F. & Evans, D. G. 2012. Familial breast cancer. *Clinical genetics* **82**: 105-114.
- Landgren, O., Goldin, L. R., Kristinsson, S. Y., Helgadottir, E. A., Samuelsson, J. & Bjorkholm, M. 2008. Increased risks of polycythemia vera, essential thrombocythemia, and myelofibrosis among 24,577 first-degree relatives of 11,039 patients with myeloproliferative neoplasms in Sweden. *Blood* **112**: 2199-2204.
- Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczy, J., LeVine, R., McEwan, P., McKernan, K., Meldrim, J., Mesirov, J. P., Miranda, C., Morris, W., Naylor, J., Raymond, C., Rosetti, M., Santos, R., Sheridan, A., Sougnez, C., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Wyman, D., Rogers, J., Sulston, J., Ainscough, R., Beck, S., Bentley, D., Burton, J., Clee, C., Carter, N., Coulson, A., Deadman, R., Deloukas, P., Dunham, A., Dunham, I., Durbin, R., French, L., Grafham, D., Gregory, S., Hubbard, T., Humphray, S., Hunt, A., Jones, M., Lloyd, C., McMurray, A., Matthews, L., Mercer, S., Milne, S., Mullikin, J. C., Mungall, A., Plumb, R., Ross, M., Shownkeen, R., Sims, S., Waterston, R. H., Wilson, R. K., Hillier, L. W., McPherson, J. D., Marra, M. A., Mardis, E. R., Fulton, L. A., Chinwalla, A. T., Pepin, K. H., Gish, W. R., Chissoe, S. L., Wendl, M. C., Delehaunty, K. D., Miner, T. L., Delehaunty, A., Kramer, J. B., Cook, L. L., Fulton, R. S., Johnson, D. L., Minx, P. J., Clifton, S. W., Hawkins, T., Branscomb, E., Predki, P., Richardson, P., Wenning, S., Slezak, T., Doggett, N., Cheng, J. F., Olsen, A., Lucas, S., Elkin, C., Uberbacher, E., Frazier, M., Gibbs, R. A., Muzny, D. M., Scherer, S. E., Bouck, J. B., Sodergren, E. J., Worley, K. C., Rives, C. M., Gorrell, J. H., Metzker, M. L., Naylor, S. L., Kucherlapati, R. S., Nelson, D. L., Weinstock, G. M., Sakaki, Y., Fujiyama, A., Hattori, M., Yada, T., Toyoda, A., Itoh, T., Kawagoe, C., Watanabe, H., Totoki, Y., Taylor, T., Weissenbach, J., Heilig, R., Saurin, W., Artiguenave, F., Brothier, P., Bruls, T., Pelletier, E., Robert, C., Wincker, P., Smith, D. R., Doucette-Stamm, L., Rubenfield, M., Weinstock, K., Lee, H. M., Dubois, J., Rosenthal, A., Platzer, M., Nyakatura, G., Taudien, S., Rump, A., Yang, H., Yu, J., Wang, J., Huang, G., Gu, J., Hood, L., Rowen, L., Madan, A., Qin, S., Davis, R. W., Federspiel, N. A., Abola, A. P., Proctor, M. J., Myers, R. M., Schmutz, J., Dickson, M., Grimwood, J., Cox, D. R., Olson, M. V., Kaul, R., Raymond, C., Shimizu, N., Kawasaki, K., Minoshima, S., Evans, G. A., Athanasiou, M., Schultz, R., Roe, B. A., Chen, F., Pan, H., Ramser, J., Lehrach, H., Reinhardt, R., McCombie, W. R., de la Bastide, M., Dedhia, N., Blocker, H., Hornischer, K., Nordsiek, G., Agarwala, R., Aravind, L., Bailey, J. A., Bateman, A., Batzoglu, S., Birney, E., Bork, P., Brown, D. G., Burge, C. B., Cerutti, L., Chen, H. C., Church, D., Clamp, M., Copley, R. R., Doerks, T., Eddy, S. R., Eichler, E. E., Furey, T. S., Galagan, J., Gilbert, J. G., Harmon, C., Hayashizaki, Y., Haussler, D., Hermjakob, H., Hokamp, K., Jang, W., Johnson, L. S., Jones, T. A., Kasif, S., Kasprzyk, A., Kennedy, S., Kent, W. J., Kitts, P., Koonin, E. V., Korf, I., Kulp, D., Lancet, D., Lowe, T. M., McLysaght, A., Mikkelsen, T., Moran, J. V., Mulder, N., Pollara, V. J., Ponting, C. P., Schuler, G., Schultz, J., Slater, G., Smit, A. F., Stupka, E., Szustakowski, J., Thierry-Mieg, D., Thierry-Mieg, J., Wagner, L., Wallis, J., Wheeler, R., Williams, A., Wolf, Y. I., Wolfe, K. H., Yang, S. P., Yeh, R. F., Collins, F., Guyer, M. S., Peterson, J., Felsenfeld, A., Wetterstrand, K. A., Patrinos, A., Morgan, M. J., de Jong, P., Catanese, J. J., Osoegawa, K., Shizuya, H., Choi, S., Chen, Y. J. & International Human Genome Sequencing Consortium. 2001. Initial sequencing and analysis of the human genome. *Nature* **409**: 860-921.
- Larjavaara, S., Haapasalo, H., Sankila, R., Helen, P. & Auvinen, A. 2008. Is the incidence of meningiomas underestimated? A regional survey. *British journal of cancer* **99**: 182-184.
- Lauper, J. M., Krause, A., Vaughan, T. L. & Monnat, R. J., Jr. 2013. Spectrum and risk of neoplasia in Werner syndrome: a systematic review. *PloS one* **8**: e59709.

- Laurendeau, I., Ferrer, M., Garrido, D., D'Haene, N., Ciavarelli, P., Basso, A., Vidaud, M., Bieche, I., Salmon, I. & Szijan, I. 2010. Gene expression profiling of the hedgehog signaling pathway in human meningiomas. *Molecular medicine (Cambridge, Mass.)* 16: 262-270.
- Lee, A. I. & LaCasce, A. S. 2009. Nodular lymphocyte predominant Hodgkin lymphoma. *The oncologist* 14: 739-751.
- Lessard, C. J., Li, H., Adrianto, I., Ice, J. A., Rasmussen, A., Grundahl, K. M., Kelly, J. A., Dozmorov, M. G., Miceli-Richard, C., Bowman, S., Lester, S., Eriksson, P., Eloranta, M. L., Brun, J. G., Goransson, L. G., Harboe, E., Guthridge, J. M., Kaufman, K. M., Kvarnstrom, M., Jazebi, H., Cunninghame Graham, D. S., Grandits, M. E., Nazmul-Hossain, A. N., Patel, K., Adler, A. J., Maier-Moore, J. S., Farris, A. D., Brennan, M. T., Lessard, J. A., Chodosh, J., Gopalakrishnan, R., Hefner, K. S., Houston, G. D., Huang, A. J., Hughes, P. J., Lewis, D. M., Radfar, L., Rohrer, M. D., Stone, D. U., Wren, J. D., Vyse, T. J., Gaffney, P. M., James, J. A., Omdal, R., Wahren-Herlenius, M., Illei, G. G., Witte, T., Jonsson, R., Rischmueller, M., Ronnblom, L., Nordmark, G., Ng, W. F., UK Primary Sjogren's Syndrome Registry, Mariette, X., Anaya, J. M., Rhodus, N. L., Segal, B. M., Scofield, R. H., Montgomery, C. G., Harley, J. B. & Sivils, K. L. 2013. Variants at multiple loci implicated in both innate and adaptive immune responses are associated with Sjogren's syndrome. *Nature genetics* 45: 1284-1292.
- Li, H. & Durbin, R. 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics (Oxford, England)* 26: 589-595.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R. & 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics (Oxford, England)* 25: 2078-2079.
- Li, Y., Schwab, C., Ryan, S. L., Papaemmanuil, E., Robinson, H. M., Jacobs, P., Moorman, A. V., Dyer, S., Borrow, J., Griffiths, M., Heerema, N. A., Carroll, A. J., Talley, P., Bown, N., Telford, N., Ross, F. M., Gaunt, L., McNally, R. J., Young, B. D., Sinclair, P., Rand, V., Teixeira, M. R., Joseph, O., Robinson, B., Maddison, M., Dastugue, N., Vandenberghe, P., Haferlach, C., Stephens, P. J., Cheng, J., Van Loo, P., Stratton, M. R., Campbell, P. J. & Harrison, C. J. 2014. Constitutional and somatic rearrangement of chromosome 21 in acute lymphoblastic leukaemia. *Nature* 508: 98-102.
- Liang, J., Lin, C., Hu, F., Wang, F., Zhu, L., Yao, X., Wang, Y. & Zhao, Y. 2013. APC polymorphisms and the risk of colorectal neoplasia: a HuGE review and meta-analysis. *American Journal of Epidemiology* 177: 1169-1179.
- Lichtenstein, P., Holm, N. V., Verkasalo, P. K., Iliadou, A., Kaprio, J., Koskenvuo, M., Pukkala, E., Skytthe, A. & Hemminki, K. 2000. Environmental and heritable factors in the causation of cancer--analyses of cohorts of twins from Sweden, Denmark, and Finland. *The New England journal of medicine* 343: 78-85.
- Lifson, A. R., Darrow, W. W., Hessol, N. A., O'Malley, P. M., Barnhart, J. L., Jaffe, H. W. & Rutherford, G. W. 1990. Kaposi's sarcoma in a cohort of homosexual and bisexual men. Epidemiology and analysis for cofactors. *American Journal of Epidemiology* 131: 221-231.
- Lim, E. T., Wurtz, P., Havulinna, A. S., Palta, P., Tukiainen, T., Rehnstrom, K., Esko, T., Magi, R., Inouye, M., Lappalainen, T., Chan, Y., Salem, R. M., Lek, M., Flannick, J., Sim, X., Manning, A., Ladenvall, C., Bumpstead, S., Hamalainen, E., Aalto, K., Maksimow, M., Salmi, M., Blankenberg, S., Ardissino, D., Shah, S., Horne, B., McPherson, R., Hovingh, G. K., Reilly, M. P., Watkins, H., Goel, A., Farrall, M., Girelli, D., Reiner, A. P., Stitzel, N. O., Kathiresan, S., Gabriel, S., Barrett, J. C., Lehtimaki, T., Laakso, M., Groop, L., Kaprio, J., Perola, M., McCarthy, M. I., Boehnke, M., Altshuler, D. M., Lindgren, C. M., Hirschhorn, J. N., Metspalu, A., Freimer, N. B., Zeller, T., Jalkanen, S., Koskinen, S., Raitakari, O., Durbin, R., MacArthur, D. G., Salomaa, V., Ripatti, S., Daly, M. J., Palotie, A. & Sequencing Initiative Suomi (SISu) Project. 2014. Distribution and Medical Impact of Loss-of-Function Variants in the Finnish Founder Population. *PLoS genetics* 10: e1004494.
- Liu, R., Paxton, W. A., Choe, S., Ceradini, D., Martin, S. R., Horuk, R., MacDonald, M. E., Stuhlmann, H., Koup, R. A. & Landau, N. R. 1996. Homozygous defect in HIV-1 coreceptor accounts for resistance of some multiply-exposed individuals to HIV-1 infection. *Cell* 86: 367-377.
- Lloyd, S. K. & Evans, D. G. 2013. Neurofibromatosis type 2 (NF2): diagnosis and management. *Handbook of clinical neurology* 115: 957-967.
- Locatelli, D., Bottoni, A., Uggetti, C. & Gozzoli, L. 1987. Multiple meningiomas evaluated by computed tomography. *Neurochirurgia* 30: 8-10.
- Lo Muzio, L. 2008. Nevoid basal cell carcinoma syndrome (Gorlin syndrome). *Orphanet journal of rare diseases* 3: 32-1172-3-32.
- Louis, D. N., Ohgaki, H., Wiestler, O. D. & Cavenee, W. K. (Editors). 2007. WHO classification of tumours of the central nervous system, fourth edition. IARC, Lyon, France.

- Ma, T., Van Tine, B. A., Wei, Y., Garrett, M. D., Nelson, D., Adams, P. D., Wang, J., Qin, J., Chow, L. T. & Harper, J. W. 2000. Cell cycle-regulated phosphorylation of p220(NPAT) by cyclin E/Cdk2 in Cajal bodies promotes histone gene transcription. *Genes & development* 14: 2298-2313.
- Malkin, D., Li, F. P., Strong, L. C., Fraumeni, J. F., Jr, Nelson, C. E., Kim, D. H., Kassel, J., Gryka, M. A., Bischoff, F. Z. & Tainsky, M. A. 1990. Germ line p53 mutations in a familial syndrome of breast cancer, sarcomas, and other neoplasms. *Science* 250: 1233-1238.
- Mancuso, R., Brambilla, L., Agostini, S., Biffi, R., Hernis, A., Guerini, F. R., Agliardi, C., Tournalaki, A., Bellinva, M. & Clerici, M. 2011. Intrafamilial transmission of Kaposi's sarcoma-associated herpesvirus and seronegative infection in family members of classic Kaposi's sarcoma patients. *The Journal of general virology* 92: 744-751.
- Mardis, E. R. 2013. Next-generation sequencing platforms. *Annual review of analytical chemistry (Palo Alto, Calif.)* 6: 287-303.
- Mardis, E. R. 2008. Next-generation DNA sequencing methods. *Annual review of genomics and human genetics* 9: 387-402.
- Mardis, E. R., Ding, L., Dooling, D. J., Larson, D. E., McLellan, M. D., Chen, K., Koboldt, D. C., Fulton, R. S., Delehaunty, K. D., McGrath, S. D., Fulton, L. A., Locke, D. P., Magrini, V. J., Abbott, R. M., Vickery, T. L., Reed, J. S., Robinson, J. S., Wylie, T., Smith, S. M., Carmichael, L., Eldred, J. M., Harris, C. C., Walker, J., Peck, J. B., Du, F., Dukes, A. F., Sanderson, G. E., Brummett, A. M., Clark, E., McMichael, J. F., Meyer, R. J., Schindler, J. K., Pohl, C. S., Wallis, J. W., Shi, X., Lin, L., Schmidt, H., Tang, Y., Haipek, C., Wiechert, M. E., Ivy, J. V., Kalicki, J., Elliott, G., Ries, R. E., Payton, J. E., Westervelt, P., Tomasson, M. H., Watson, M. A., Baty, J., Heath, S., Shannon, W. D., Nagarajan, R., Link, D. C., Walter, M. J., Graubert, T. A., DiPersio, J. F., Wilson, R. K. & Ley, T. J. 2009. Recurring mutations found by sequencing an acute myeloid leukemia genome. *The New England journal of medicine* 361: 1058-1066.
- Marsh, D. & Zori, R. 2002. Genetic insights into familial cancers-- update and recent discoveries. *Cancer letters* 181: 125-164.
- Mason, D. Y., Banks, P. M., Chan, J., Cleary, M. L., Delsol, G., de Wolf Peeters, C., Falini, B., Gatter, K., Grogan, T. M. & Harris, N. L. 1994. Nodular lymphocyte predominance Hodgkin's disease. A distinct clinicopathological entity. *The American Journal of Surgical Pathology* 18: 526-530.
- Maston, G. A., Evans, S. K. & Green, M. R. 2006. Transcriptional regulatory elements in the human genome. *Annual review of genomics and human genetics* 7: 29-59.
- Matikainen, M. P., Sankila, R., Schleutker, J., Kallioniemi, O. P. & Pukkala, E. 2000. Nationwide cancer family ascertainment using Finnish Cancer Registry data on family names and places of birth for 35,761 prostate cancer patients. *International journal of cancer* 88: 307-312.
- Mayama, S., Cuevas, L. E., Sheldon, J., Omar, O. H., Smith, D. H., Okong, P., Silvel, B., Hart, C. A. & Schulz, T. F. 1998. Prevalence and transmission of Kaposi's sarcoma-associated herpesvirus (human herpesvirus 8) in Ugandan children and adolescents. *International journal of cancer* 77: 817-820.
- Mbulaiteye, S. M. & Engels, E. A. 2006. Kaposi's sarcoma risk among transplant recipients in the United States (1993-2003). *International journal of cancer* 119: 2685-2691.
- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernysky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M. & DePristo, M. A. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome research* 20: 1297-1303.
- Medina, R., van der Deen, M., Miele-Chamberland, A., Xie, R. L., van Wijnen, A. J., Stein, J. L. & Stein, G. S. 2007. The HNF-P/p220NPAT cell cycle signaling pathway controls nonhistone target genes. *Cancer research* 67: 10334-10342.
- Mehine, M., Kaasinen, E., Mäkinen, N., Katainen, R., Kämpjarvi, K., Pitkänen, E., Heinonen, H. R., Bützow, R., Kilpivaara, O., Kuosmanen, A., Ristolainen, H., Gentile, M., Sjöberg, J., Vahteristo, P. & Aaltonen, L. A. 2013. Characterization of uterine leiomyomas by whole-genome sequencing. *The New England journal of medicine* 369: 43-53.
- Meijers-Heijboer, H., van den Ouweland, A., Klijn, J., Wasielewski, M., de Snoo, A., Oldenburg, R., Hollestelle, A., Houben, M., Crepin, E., van Veghel-Plandsoen, M., Elstrodt, F., van Duijn, C., Bartels, C., Meijers, C., Schutte, M., McGuffog, L., Thompson, D., Easton, D., Sodha, N., Seal, S., Barfoot, R., Mangion, J., Chang-Claude, J., Eccles, D., Eeles, R., Evans, D. G., Houlston, R., Murday, V., Narod, S., Peretz, T., Peto, J., Phelan, C., Zhang, H. X., Szabo, C., Devilee, P., Goldgar, D., Futreal, P. A., Nathanson, K. L., Weber, B., Rahman, N., Stratton, M. R. & CHEK2-Breast Cancer Consortium. 2002. Low-penetrance susceptibility to breast cancer due to CHEK2(*)1100delC in noncarriers of BRCA1 or BRCA2 mutations. *Nature genetics* 31: 55-59.
- Melean, G., Velasco, A., Hernandez-Imaz, E., Rodriguez-Alvarez, F. J., Martin, Y., Valero, A. & Hernandez-Chico, C. 2012. RNA-based analysis of two SMARCB1 mutations associated with familial schwannomatosis with meningiomas. *Neurogenetics* 13: 267-274.

- Merchant, M., Vajdos, F. F., Ultsch, M., Maun, H. R., Wendt, U., Cannon, J., Desmarais, W., Lazarus, R. A., de Vos, A. M. & de Sauvage, F. J. 2004. Suppressor of fused regulates Gli activity through a dual binding mechanism. *Molecular and cellular biology* 24: 8627-8641.
- Mesri, E. A., Cesarman, E. & Boshoff, C. 2010. Kaposi's sarcoma and its associated herpesvirus. *Nature reviews.Cancer* 10: 707-719.
- Mesri, E. A., Feitelson, M. A. & Munger, K. 2014. Human Viral Oncogenesis: A Cancer Hallmarks Analysis. *Cell host & microbe* 15: 266-282.
- Mills, R. E., Bennett, E. A., Iskow, R. C. & Devine, S. E. 2007. Which transposable elements are active in the human genome? *Trends in genetics : TIG* 23: 183-191.
- Mishra, D. K., Chen, Z., Wu, Y., Sarkissyan, M., Koeffler, H. P. & Vadgama, J. V. 2010. Global methylation pattern of genes in androgen-sensitive and androgen-independent prostate cancer cells. *Molecular cancer therapeutics* 9: 33-45.
- Miyaki, M., Konishi, M., Tanaka, K., Kikuchi-Yanoshita, R., Muraoka, M., Yasuno, M., Igari, T., Koike, M., Chiba, M. & Mori, T. 1997. Germline mutation of MSH6 as the cause of hereditary nonpolyposis colorectal cancer. *Nature genetics* 17: 271-272.
- Morton, L. M., Wang, S. S., Devesa, S. S., Hartge, P., Weisenburger, D. D. & Linet, M. S. 2006. Lymphoma incidence patterns by WHO subtype in the United States, 1992-2001. *Blood* 107: 265-276.
- Morton, N. E. 1955. Sequential tests for the detection of linkage. *American Journal of Human Genetics* 7: 277-318.
- Nagy, R., Sweet, K. & Eng, C. 2004. Highly penetrant hereditary cancer syndromes. *Oncogene* 23: 6445-6470.
- Nahser, H. C., Grote, W., Lohr, E. & Gerhard, L. 1981. Multiple meningiomas. Clinical and computer tomographic observations. *Neuroradiology* 21: 259-263.
- Nichols, A. F., Ong, P. & Linn, S. 1996. Mutations specific to the xeroderma pigmentosum group E Ddb- phenotype. *The Journal of biological chemistry* 271: 24317-24320.
- Nicolaides, N. C., Papadopoulos, N., Liu, B., Wei, Y. F., Carter, K. C., Ruben, S. M., Rosen, C. A., Haseltine, W. A., Fleischmann, R. D. & Fraser, C. M. 1994. Mutations of two PMS homologues in hereditary nonpolyposis colon cancer. *Nature* 371: 75-80.
- Niemann, S. & Muller, U. 2000. Mutations in SDHC cause autosomal dominant paraganglioma, type 3. *Nature genetics* 26: 268-270.
- Nik-Zainal, S., Alexandrov, L. B., Wedge, D. C., Van Loo, P., Greenman, C. D., Raine, K., Jones, D., Hinton, J., Marshall, J., Stebbings, L. A., Menzies, A., Martin, S., Leung, K., Chen, L., Leroy, C., Ramakrishna, M., Rance, R., Lau, K. W., Mudie, L. J., Varela, I., McBride, D. J., Bignell, G. R., Cooke, S. L., Shlien, A., Gamble, J., Whitmore, I., Maddison, M., Tarpey, P. S., Davies, H. R., Papaemmanuil, E., Stephens, P. J., McLaren, S., Butler, A. P., Teague, J. W., Jonsson, G., Garber, J. E., Silver, D., Miron, P., Fatima, A., Boyault, S., Langerod, A., Tutt, A., Martens, J. W., Aparicio, S. A., Borg, A., Salomon, A. V., Thomas, G., Borresen-Dale, A. L., Richardson, A. L., Neuberger, M. S., Futreal, P. A., Campbell, P. J., Stratton, M. R. & Breast Cancer Working Group of the International Cancer Genome Consortium. 2012. Mutational processes molding the genomes of 21 breast cancers. *Cell* 149: 979-993.
- Nishida, T., Hirota, S., Taniguchi, M., Hashimoto, K., Isozaki, K., Nakamura, H., Kanakura, Y., Tanaka, T., Takabayashi, A., Matsuda, H. & Kitamura, Y. 1998. Familial gastrointestinal stromal tumours with germline mutation of the KIT gene. *Nature genetics* 19: 323-324.
- Niu, D. K. & Jiang, L. 2013. Can ENCODE tell us how much junk DNA we carry in our genome? *Biochemical and biophysical research communications* 430: 1340-1343.
- Nogova, L., Reineke, T., Brillant, C., Sieniawski, M., Rudiger, T., Josting, A., Bredenfeld, H., Skripnitchenko, R., Muller, R. P., Muller-Hermelink, H. K., Diehl, V., Engert, A. & German Hodgkin Study Group. 2008. Lymphocyte-predominant and classical Hodgkin's lymphoma: a comprehensive analysis from the German Hodgkin Study Group. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 26: 434-439.
- Nose, V. 2011. Familial thyroid cancer: a review. *Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc* 24 Suppl 2: S19-33.
- Nowell, P. C. 1976. The clonal evolution of tumor cell populations. *Science* 194: 23-28.
- Ogoina, D., Onyemelukwe, G., Musa, B. O. & Babadoko, A. 2011. Seroprevalence and determinants of human herpes virus 8 infection in adult Nigerians with and without HIV-1 infection. *African health sciences* 11: 158-162.
- Olshen, A. B., Venkatraman, E. S., Lucito, R. & Wigler, M. 2004. Circular binary segmentation for the analysis of array-based DNA copy number data. *Biostatistics (Oxford, England)* 5: 557-572.
- Paigen, K. & Petkov, P. 2010. Mammalian recombination hot spots: properties, control and evolution. *Nature reviews.Genetics* 11: 221-233.

- Palles, C., Cazier, J. B., Howarth, K. M., Domingo, E., Jones, A. M., Broderick, P., Kemp, Z., Spain, S. L., Guarino, E., Salguero, I., Sherborne, A., Chubb, D., Carvajal-Carmona, L. G., Ma, Y., Kaur, K., Dobbins, S., Barclay, E., Gorman, M., Martin, L., Kovac, M. B., Humphray, S., CORGI Consortium, WGS500 Consortium, Lucassen, A., Holmes, C. C., Bentley, D., Donnelly, P., Taylor, J., Petridis, C., Roylance, R., Sawyer, E. J., Kerr, D. J., Clark, S., Grimes, J., Kearsey, S. E., Thomas, H. J., McVean, G., Houlston, R. S. & Tomlinson, I. 2013. Germline mutations affecting the proofreading domains of POLE and POLD1 predispose to colorectal adenomas and carcinomas. *Nature genetics* 45: 136-144.
- Papasteriades, C., Kaloterakis, A., Filiotou, A., Economidou, J., Nicolis, G., Trichopoulos, D. & Stratigos, J. 1984. Histocompatibility antigens HLA-A, -B, -DR in Greek patients with Kaposi's sarcoma. *Tissue antigens* 24: 313-315.
- Parker, M., Mohankumar, K. M., Punchihewa, C., Weinlich, R., Dalton, J. D., Li, Y., Lee, R., Tatevossian, R. G., Phoenix, T. N., Thiruvengatam, R., White, E., Tang, B., Orisme, W., Gupta, K., Rusch, M., Chen, X., Li, Y., Nagahawhatte, P., Hedlund, E., Finkelstein, D., Wu, G., Shurtleff, S., Easton, J., Boggs, K., Yergeau, D., Vadodaria, B., Mulder, H. L., Becksford, J., Gupta, P., Huether, R., Ma, J., Song, G., Gajjar, A., Merchant, T., Boop, F., Smith, A. A., Ding, L., Lu, C., Ochoa, K., Zhao, D., Fulton, R. S., Fulton, L. L., Mardis, E. R., Wilson, R. K., Downing, J. R., Green, D. R., Zhang, J., Ellison, D. W. & Gilbertson, R. J. 2014. C11orf95-RELA fusions drive oncogenic NF-kappaB signalling in ependymoma. *Nature* 506: 451-455.
- Parkin, D. M., Wabinga, H., Nambooz, S. & Wabwire-Mangen, F. 1999. AIDS-related cancers in Africa: maturation of the epidemic in Uganda. *AIDS (London, England)* 13: 2563-2570.
- Pastorino, L., Ghiorzo, P., Nasti, S., Battistuzzi, L., Cusano, R., Marzocchi, C., Garre, M. L., Clementi, M. & Scarra, G. B. 2009. Identification of a SUFU germline mutation in a family with Gorlin syndrome. *American journal of medical genetics. Part A* 149A: 1539-1543.
- Payne, S. R. & Kemp, C. J. 2005. Tumor suppressor genetics. *Carcinogenesis* 26: 2031-2045.
- Penegar, S., Wood, W., Lubbe, S., Chandler, I., Broderick, P., Papaemmanuil, E., Sellick, G., Gray, R., Peto, J. & Houlston, R. 2007. National study of colorectal cancer genetics. *British journal of cancer* 97: 1305-1309.
- Peutz, J. 1921. *Very remarkable case of familial polyposis of mucous membrane of intestinal tract and nasopharynx accompanied by peculiar pigmentations of skin and mucous membrane.* 10: 134-146.
- Pfeifer, G. P. 2010. Environmental exposures and mutational patterns of cancer genomes. *Genome medicine* 2: 54.
- Picard, C., Mellouli, F., Duprez, R., Chedeville, G., Neven, B., Fraitag, S., Delaunay, J., Le Deist, F., Fischer, A., Blanche, S., Bodemer, C., Gessain, A., Casanova, J. L. & Bejaoui, M. 2006. Kaposi's sarcoma in a child with Wiskott-Aldrich syndrome. *European journal of pediatrics* 165: 453-457.
- Pickrell, J. K., Gaffney, D. J., Gilad, Y. & Pritchard, J. K. 2011. False positive peaks in ChIP-seq and other sequencing-based functional assays caused by unannotated high copy number regions. *Bioinformatics (Oxford, England)* 27: 2144-2146.
- Potter, L. M., Linet, M., Blair, A., Dick, F., Burmeister, L. F., Gibson, R., Schuman, L. M. & Fraumeni, J. F., Jr. 1991. Familial cancers associated with subtypes of leukemia and non-Hodgkin's lymphoma. *Leukemia research* 15: 305-314.
- Pribila, J. T., Ronan, S. M. & Trobe, J. D. 2008. Multiple intracranial meningiomas causing papilledema and visual loss in a patient with nevoid Basal cell carcinoma syndrome. *Journal of neuro-ophthalmology : the official journal of the North American Neuro-Ophthalmology Society* 28: 41-46.
- Pukkala, E. & Patama, T. 2010. Small-area based map animations of cancer incidence in Finland, 1953–2008. Accessed 16 July 2012:
- Pukkala, E. 2011. Biobanks and registers in epidemiologic research on cancer. *Methods in molecular biology (Clifton, N.J.)* 675: 127-164.
- Pulst, S. M. 1999. Genetic linkage analysis. *Archives of Neurology* 56: 667-672.
- Pykaläinen, M., Kinos, R., Valkonen, S., Rydman, P., Kilpeläinen, M., Laitinen, L. A., Karjalainen, J., Nieminen, M., Hurme, M., Kere, J., Laitinen, T. & Laheesmaa, R. 2005. Association analysis of common variants of STAT6, GATA3, and STAT4 to asthma and high serum IgE phenotypes. *The Journal of allergy and clinical immunology* 115: 80-87.
- Radovanovic, Z., Markovic-Denic, L. & Jankovic, S. 1994. Cancer mortality of family members of patients with chronic lymphocytic leukemia. *European journal of epidemiology* 10: 211-213.
- Rafnar, T., Gudbjartsson, D. F., Sulem, P., Jonasdottir, A., Sigurdsson, A., Jonasdottir, A., Besenbacher, S., Lundin, P., Stacey, S. N., Gudmundsson, J., Magnusson, O. T., le Roux, L., Orlygsdottir, G., Helgadottir, H. T., Johannsdottir, H., Gylfason, A., Tryggvadottir, L., Jonasson, J. G., de Juan, A., Ortega, E., Ramon-Cajal, J. M., Garcia-Prats, M. D., Mayordomo, C., Panadero, A., Rivera, F., Aben, K. K., van Altena, A. M., Massuger, L. F., Aavikko, M., Kujala, P. M., Staff, S., Aaltonen, L. A., Olafsdottir, K., Bjornsson, J., Kong, A., Salvarsdottir, A., Saemundsson, H., Olafsson, K., Benediksdottir, K. R., Gulcher, J., Masson, G., Kiemenev, L. A., Mayordomo,

- J. I., Thorsteinsdottir, U. & Stefansson, K. 2011. Mutations in BRIP1 confer high risk of ovarian cancer. *Nature genetics* 43: 1104-1107.
- Rahman, N. 2014. Realizing the promise of cancer predisposition genes. *Nature* 505: 302-308.
- Rahman, N., Seal, S., Thompson, D., Kelly, P., Renwick, A., Elliott, A., Reid, S., Spanova, K., Barfoot, R., Chagtai, T., Jayatilake, H., McGuffog, L., Hanks, S., Evans, D. G., Eccles, D., Breast Cancer Susceptibility Collaboration (UK), Easton, D. F. & Stratton, M. R. 2007. PALB2, which encodes a BRCA2-interacting protein, is a breast cancer susceptibility gene. *Nature genetics* 39: 165-167.
- Reifenberger, J., Wolter, M., Weber, R. G., Megahed, M., Ruzicka, T., Lichter, P. & Reifenberger, G. 1998. Missense mutations in SMOH in sporadic basal cell carcinomas of the skin and primitive neuroectodermal tumors of the central nervous system. *Cancer research* 58: 1798-1803.
- Remmers, E. F., Plenge, R. M., Lee, A. T., Graham, R. R., Hom, G., Behrens, T. W., de Bakker, P. I., Le, J. M., Lee, H. S., Batliwalla, F., Li, W., Masters, S. L., Booty, M. G., Carulli, J. P., Padyukov, L., Alfredsson, L., Klareskog, L., Chen, W. V., Amos, C. I., Criswell, L. A., Seldin, M. F., Kastner, D. L. & Gregersen, P. K. 2007. STAT4 and the risk of rheumatoid arthritis and systemic lupus erythematosus. *The New England journal of medicine* 357: 977-986.
- Renwick, A., Thompson, D., Seal, S., Kelly, P., Chagtai, T., Ahmed, M., North, B., Jayatilake, H., Barfoot, R., Spanova, K., McGuffog, L., Evans, D. G., Eccles, D., Breast Cancer Susceptibility Collaboration (UK), Easton, D. F., Stratton, M. R. & Rahman, N. 2006. ATM mutations that cause ataxia-telangiectasia are breast cancer susceptibility alleles. *Nature genetics* 38: 873-875.
- Reya, T., Morrison, S. J., Clarke, M. F. & Weissman, I. L. 2001. Stem cells, cancer, and cancer stem cells. *Nature* 414: 105-111.
- Risch, N. & Merikangas, K. 1996. The future of genetic studies of complex human diseases. *Science* 273: 1516-1517.
- Roberts, W. M., Douglass, E. C., Peiper, S. C., Houghton, P. J. & Look, A. T. 1989. Amplification of the gli gene in childhood sarcomas. *Cancer research* 49: 5407-5413.
- Robinson, S. & Cohen, A. R. 2000. Cowden disease and Lhermitte-Duclos disease: characterization of a new phakomatosis. *Neurosurgery* 46: 371-383.
- Rudin, C. M. 2012. Vismodegib. *Clinical cancer research : an official journal of the American Association for Cancer Research* 18: 3218-3222.
- Ruel, L. & Therond, P. P. 2009. Variations in Hedgehog signaling: divergence and perpetuation in Sufu regulation of Gli. *Genes & development* 23: 1843-1848.
- Saarinén, S., Kaasinen, E., Karjalainen-Lindsberg, M. L., Vesanen, K., Aavikko, M., Katainen, R., Taskinen, M., Kytölä, S., Leppä, S., Hietala, M., Vahteristo, P. & Aaltonen, L. A. 2013. Primary mediastinal large B-cell lymphoma segregating in a family: exome sequencing identifies MLL as a candidate predisposition gene. *Blood* 121: 3428-3430.
- Saarinén, S., Pukkala, E., Vahteristo, P., Mäkinen, M. J., Franssila, K. & Aaltonen, L. A. 2013. High familial risk in nodular lymphocyte-predominant Hodgkin lymphoma. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 31: 938-943.
- Saarinén, S., Vahteristo, P., Launonen, V., Franssila, K., Kivirikko, S., Lehtonen, R., Bain, B. J., Bauduer, F., Unal, A., Aaltonen, L. A. & Aittomäki, K. 2011. Analysis of KLHDC8B in familial nodular lymphocyte predominant Hodgkin lymphoma. *British journal of haematology* 154: 413-415.
- Samson, M., Libert, F., Doranz, B. J., Rucker, J., Liesnard, C., Farber, C. M., Saragosti, S., Lapoumeroulie, C., Cogniaux, J., Forceille, C., Muyldermans, G., Verhofstede, C., Burtonboy, G., Georges, M., Imai, T., Rana, S., Yi, Y., Smyth, R. J., Collman, R. G., Doms, R. W., Vassart, G. & Parmentier, M. 1996. Resistance to HIV-1 infection in caucasian individuals bearing mutant alleles of the CCR-5 chemokine receptor gene. *Nature* 382: 722-725.
- Santarosa, M. & Ashworth, A. 2004. Haploinsufficiency for tumour suppressor genes: when you don't need to go all the way. *Biochimica et biophysica acta* 1654: 105-122.
- Savitsky, K., Bar-Shira, A., Gilad, S., Rotman, G., Ziv, Y., Vanagaite, L., Tagle, D. A., Smith, S., Uziel, T., Sfez, S., Ashkenazi, M., Pecker, I., Frydman, M., Harnik, R., Patanjali, S. R., Simmons, A., Clines, G. A., Sartiel, A., Gatti, R. A., Chessa, L., Sanal, O., Lavin, M. F., Jaspers, N. G., Taylor, A. M., Arlett, C. F., Miki, T., Weissman, S. M., Lovett, M., Collins, F. S. & Shiloh, Y. 1995a. A single ataxia telangiectasia gene with a product similar to PI-3 kinase. *Science* 268: 1749-1753.
- Savitsky, K., Sfez, S., Tagle, D. A., Ziv, Y., Sartiel, A., Collins, F. S., Shiloh, Y. & Rotman, G. 1995b. The complete sequence of the coding region of the ATM gene reveals similarity to cell cycle regulators in different species. *Human molecular genetics* 4: 2025-2032.
- Seal, S., Thompson, D., Renwick, A., Elliott, A., Kelly, P., Barfoot, R., Chagtai, T., Jayatilake, H., Ahmed, M., Spanova, K., North, B., McGuffog, L., Evans, D. G., Eccles, D., Breast Cancer Susceptibility Collaboration

- (UK), Easton, D. F., Stratton, M. R. & Rahman, N. 2006. Truncating mutations in the Fanconi anemia J gene BRIP1 are low-penetrance breast cancer susceptibility alleles. *Nature genetics* 38: 1239-1241.
- Sherr, C. J. 2004. Principles of tumor suppression. *Cell* 116: 235-246.
- Shi, Q., Han, Y. & Jiang, J. 2014. Suppressor of fused impedes Ci/Gli nuclear import by opposing Trn/Kapbeta2 in Hedgehog signaling. *Journal of cell science* 127: 1092-1103.
- Sijbers, A. M., de Laat, W. L., Ariza, R. R., Biggerstaff, M., Wei, Y. F., Moggs, J. G., Carter, K. C., Shell, B. K., Evans, E., de Jong, M. C., Rademakers, S., de Rooij, J., Jaspers, N. G., Hoeijmakers, J. H. & Wood, R. D. 1996. Xeroderma pigmentosum group F caused by a defect in a structure-specific DNA repair endonuclease. *Cell* 86: 811-822.
- Smith, L. B. 2010. Nodular lymphocyte predominant Hodgkin lymphoma: diagnostic pearls and pitfalls. *Archives of Pathology & Laboratory Medicine* 134: 1434-1439.
- Smith, M. J., O'Sullivan, J., Bhaskar, S. S., Hadfield, K. D., Poke, G., Caird, J., Sharif, S., Eccles, D., Fitzpatrick, D., Rawluk, D., du Plessis, D., Newman, W. G. & Evans, D. G. 2013. Loss-of-function mutations in SMARCE1 cause an inherited disorder of multiple spinal meningiomas. *Nature genetics* 45: 295-298.
- Smith, M. J., Wallace, A. J., Bennett, C., Hasselblatt, M., Elert-Dobkowska, E., Evans, L. T., Hickey, W. F., van Hoff, J., Bauer, D., Lee, A., Hevner, R. F., Beetz, C., du Plessis, D., Kilday, J. P., Newman, W. G. & Evans, D. G. 2014. Germline SMARCE1 mutations predispose to both spinal and cranial clear cell meningiomas. *The Journal of pathology*
- Smith, S. M., Anastasi, J., Cohen, K. S. & Godley, L. A. 2010. The impact of MYC expression in lymphoma biology: beyond Burkitt lymphoma. *Blood cells, molecules & diseases* 45: 317-323.
- Sneller, M. C., Wang, J., Dale, J. K., Strober, W., Middleton, L. A., Choi, Y., Fleisher, T. A., Lim, M. S., Jaffe, E. S., Puck, J. M., Lenardo, M. J. & Straus, S. E. 1997. Clinical, immunologic, and genetic features of an autoimmune lymphoproliferative syndrome associated with abnormal lymphocyte apoptosis. *Blood* 89: 1341-1348.
- Sobota, A., Pena, M., Santi, M. & Ali Ahmed, A. 2007. Undifferentiated sinonasal carcinoma in a patient with nevoid basal cell carcinoma syndrome. *International journal of surgical pathology* 15: 303-306.
- Solimini, N. L., Xu, Q., Mermel, C. H., Liang, A. C., Schlabach, M. R., Luo, J., Burrows, A. E., Anselmo, A. N., Bredemeyer, A. L., Li, M. Z., Beroukhir, R., Meyerson, M. & Elledge, S. J. 2012. Recurrent hemizygous deletions in cancers may optimize proliferative potential. *Science* 337: 104-109.
- Steffen, J., Nowakowska, D., Niwinska, A., Czapczak, D., Kluska, A., Piatkowska, M., Wisniewska, A. & Paszko, Z. 2006. Germline mutations 657del5 of the NBS1 gene contribute significantly to the incidence of breast cancer in Central Poland. *International journal of cancer* 119: 472-475.
- Stephens, P. J., Greenman, C. D., Fu, B., Yang, F., Bignell, G. R., Mudie, L. J., Pleasance, E. D., Lau, K. W., Beare, D., Stebbings, L. A., McLaren, S., Lin, M. L., McBride, D. J., Varela, I., Nik-Zainal, S., Leroy, C., Jia, M., Menzies, A., Butler, A. P., Teague, J. W., Quail, M. A., Burton, J., Swerdlow, H., Carter, N. P., Morsberger, L. A., Iacobuzio-Donahue, C., Follows, G. A., Green, A. R., Flanagan, A. M., Stratton, M. R., Futreal, P. A. & Campbell, P. J. 2011. Massive genomic rearrangement acquired in a single catastrophic event during cancer development. *Cell* 144: 27-40.
- Svärd, J., Heby-Henricson, K., Persson-Lek, M., Rozell, B., Lauth, M., Bergström, A., Ericson, J., Toftgård, R. & Teglund, S. 2006. Genetic elimination of Suppressor of fused reveals an essential repressor function in the mammalian Hedgehog signaling pathway. *Developmental cell* 10: 187-197.
- Svensson, A., Tunback, P., Nordström, I., Shestakov, A., Padyukov, L. & Eriksson, K. 2012. STAT4 regulates antiviral gamma interferon responses and recurrent disease during herpes simplex virus 2 infection. *Journal of virology* 86: 9409-9415.
- Swerdlow, S. H., Campo, E., Lee Harris, N., Jaffe, E. S., Pileri, S. A., Stien, H., Thiele, J. & Vardiman, J. W. 2008. WHO Classification of Tumours of Haematopoietic and Lymphoid Tissues, 4th edition. 323-325.
- Taipale, J., Chen, J. K., Cooper, M. K., Wang, B., Mann, R. K., Milenkovic, L., Scott, M. P. & Beachy, P. A. 2000. Effects of oncogenic mutations in Smoothened and Patched can be reversed by cyclopamine. *Nature* 406: 1005-1009.
- Takeda, K. & Akira, S. 2000. STAT family of transcription factors in cytokine-mediated biological responses. *Cytokine & growth factor reviews* 11: 199-207.
- Tate, G., Li, M., Suzuki, T. & Mitsuya, T. 2003. A new germline mutation of the PTCH gene in a Japanese patient with nevoid basal cell carcinoma syndrome associated with meningioma. *Japanese journal of clinical oncology* 33: 47-50.
- Taylor, M. D., Liu, L., Raffel, C., Hui, C. C., Mainprize, T. G., Zhang, X., Agatep, R., Chiappa, S., Gao, L., Lowrance, A., Hao, A., Goldstein, A. M., Stavrou, T., Scherer, S. W., Dura, W. T., Wainwright, B., Squire, J. A., Rutka, J. T. & Hogg, D. 2002. Mutations in SUFU predispose to medulloblastoma. *Nature genetics* 31: 306-310.

- Tedeschi, R., Bidoli, E., Agren, A., Hallmans, G., Wadell, G., De Paoli, P. & Dillner, J. 2006. Epidemiology of Kaposi's Sarcoma herpesvirus (HHV8) in Vasterbotten County, Sweden. *Journal of medical virology* 78: 372-378.
- Teppo, L., Pukkala, E. & Lehtonen, M. 1994. Data quality and quality control of a population-based cancer registry. Experience in Finland. *Acta Oncologica (Stockholm, Sweden)* 33: 365-369.
- Testa, J. R., Cheung, M., Pei, J., Below, J. E., Tan, Y., Sementino, E., Cox, N. J., Dogan, A. U., Pass, H. I., Trusa, S., Hesdorffer, M., Nasu, M., Powers, A., Rivera, Z., Comertpay, S., Tanji, M., Gaudino, G., Yang, H. & Carbone, M. 2011. Germline BAP1 mutations predispose to malignant mesothelioma. *Nature genetics* 43: 1022-1025.
- Thiagalingam, S., Lengauer, C., Leach, F. S., Schutte, M., Hahn, S. A., Overhauser, J., Willson, J. K., Markowitz, S., Hamilton, S. R., Kern, S. E., Kinzler, K. W. & Vogelstein, B. 1996. Evaluation of candidate tumour suppressor genes on chromosome 18 in colorectal cancers. *Nature genetics* 13: 343-346.
- Thierfelder, W. E., van Deursen, J. M., Yamamoto, K., Tripp, R. A., Sarawar, S. R., Carson, R. T., Sangster, M. Y., Vignali, D. A., Doherty, P. C., Grosveld, G. C. & Ihle, J. N. 1996. Requirement for Stat4 in interleukin-12-mediated responses of natural killer and T cells. *Nature* 382: 171-174.
- Thomas, A., Cannon-Albright, L., Bansal, A. & Skolnick, M. H. 1999. Familial Associations between Cancer Sites. *Computers and Biomedical Research* 32: 517-529.
- Tomasetti, C., Vogelstein, B. & Parmigiani, G. 2013. Half or more of the somatic mutations in cancers of self-renewing tissues originate prior to tumor initiation. *Proceedings of the National Academy of Sciences of the United States of America* 110: 1999-2004.
- Toyoda, H., Ido, M., Hayashi, T., Gabazza, E. C., Suzuki, K., Bu, J., Tanaka, S., Nakano, T., Kamiya, H., Chipeta, J., Kisenge, R. R., Kang, J., Hori, H. & Komada, Y. 2004. Impairment of IL-12-dependent STAT4 nuclear translocation in a patient with recurrent Mycobacterium avium infection. *Journal of immunology* 172: 3905-3912.
- Ulahannan, D., Kovac, M. B., Mulholland, P. J., Cazier, J. B. & Tomlinson, I. 2013. Technical and implementation issues in using next-generation sequencing of cancers in clinical practice. *British journal of cancer* 109: 827-835.
- Unal, A., Sari, I., Deniz, K., Ozkan, M., Kontas, O., Eser, B. & Cetin, M. 2005. Familial nodular lymphocyte predominant Hodgkin lymphoma: Successful treatment with CHOP plus rituximab. *Leukemia & lymphoma* 46: 1613-1617.
- Ur Rehman, J., Burney, I. A., Al Kindi, S. & Raeburn, S. 2008. Familial lymphoma in an Omani kindred with identical class II HLA type. *Leukemia & lymphoma* 49: 1407-1410.
- Vahteristo, P., Bartkova, J., Eerola, H., Syrjakoski, K., Ojala, S., Kilpivaara, O., Tamminen, A., Kononen, J., Aittomäki, K., Heikkilä, P., Holli, K., Blomqvist, C., Bartek, J., Kallioniemi, O. P. & Nevanlinna, H. 2002. A CHEK2 genetic variant contributing to a substantial fraction of familial breast cancer. *American Journal of Human Genetics* 71: 432-438.
- Vaittinen, P. & Hemminki, K. 1999. Familial cancer risks in offspring from discordant parental cancers. *International journal of cancer* 81: 12-19.
- van Alkemade, H., de Leau, M., Dieleman, E. M., Kardaun, J. W., van Os, R., Vandertop, W. P., van Furth, W. R. & Stalpers, L. J. 2012. Impaired survival and long-term neurological problems in benign meningioma. *Neuro-oncology* 14: 658-666.
- van den Berg, A., Maggio, E., Diepstra, A., de Jong, D., van Krieken, J. & Poppema, S. 2002. Germline FAS gene mutation in a case of ALPS and NLP Hodgkin lymphoma. *Blood* 99: 1492-1494.
- van den Munckhof, P., Christiaans, I., Kenter, S. B., Baas, F. & Hulsebos, T. J. 2012. Germline SMARCB1 mutation predisposes to multiple meningiomas and schwannomas with preferential location of cranial meningiomas at the falx cerebri. *Neurogenetics* 13: 1-7.
- van Meurs, H. S., Dieles, J. J. & Stel, H. V. 2012. A uterine leiomyoma in which a leiomyosarcoma with osteoclast-like giant cells and a metastasis of a ductal breast carcinoma are present. *Annals of Diagnostic Pathology* 16: 67-70.
- Varjosalo, M., Li, S. P. & Taipale, J. 2006. Divergence of hedgehog signal transduction mechanism between Drosophila and mammals. *Developmental cell* 10: 177-186.
- Varjosalo, M. & Taipale, J. 2008. Hedgehog: functions and mechanisms. *Genes & development* 22: 2454-2472.
- Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., Smith, H. O., Yandell, M., Evans, C. A., Holt, R. A., Gocayne, J. D., Amanatides, P., Ballew, R. M., Huson, D. H., Wortman, J. R., Zhang, Q., Kodira, C. D., Zheng, X. H., Chen, L., Skupski, M., Subramanian, G., Thomas, P. D., Zhang, J., Gabor Miklos, G. L., Nelson, C., Broder, S., Clark, A. G., Nadeau, J., McKusick, V. A., Zinder, N., Levine, A. J., Roberts, R. J., Simon, M., Slayman, C., Hunkapiller, M., Bolanos, R., Delcher, A., Dew, I., Fasulo, D., Flanigan, M., Florea, L., Halpern, A., Hannenhalli, S., Kravitz, S., Levy, S., Mobarry, C., Reinert, K., Remington, K., Abu-Threideh, J., Beasley, E., Biddick, K., Bonazzi, V., Brandon, R., Cargill, M., Chandramouliswaran, I., Charlab, R.,

- Chaturvedi, K., Deng, Z., Di Francesco, V., Dunn, P., Eilbeck, K., Evangelista, C., Gabrielian, A. E., Gan, W., Ge, W., Gong, F., Gu, Z., Guan, P., Heiman, T. J., Higgins, M. E., Ji, R. R., Ke, Z., Ketchum, K. A., Lai, Z., Lei, Y., Li, Z., Li, J., Liang, Y., Lin, X., Lu, F., Merkulov, G. V., Milshina, N., Moore, H. M., Naik, A. K., Narayan, V. A., Neelam, B., Nusskern, D., Rusch, D. B., Salzberg, S., Shao, W., Shue, B., Sun, J., Wang, Z., Wang, A., Wang, X., Wang, J., Wei, M., Wides, R., Xiao, C., Yan, C., Yao, A., Ye, J., Zhan, M., Zhang, W., Zhang, H., Zhao, Q., Zheng, L., Zhong, F., Zhong, W., Zhu, S., Zhao, S., Gilbert, D., Baumhueter, S., Spier, G., Carter, C., Cravchik, A., Woodage, T., Ali, F., An, H., Awe, A., Baldwin, D., Baden, H., Barnstead, M., Barrow, I., Beeson, K., Busam, D., Carver, A., Center, A., Cheng, M. L., Curry, L., Danaher, S., Davenport, L., Desilets, R., Dietz, S., Dodson, K., Doup, L., Ferreira, S., Garg, N., Gluecksmann, A., Hart, B., Haynes, J., Haynes, C., Heiner, C., Hladun, S., Hostin, D., Houck, J., Howland, T., Ibegwam, C., Johnson, J., Kalush, F., Kline, L., Koduru, S., Love, A., Mann, F., May, D., McCawley, S., McIntosh, T., McMullen, I., Moy, M., Moy, L., Murphy, B., Nelson, K., Pfannkoch, C., Pratts, E., Puri, V., Qureshi, H., Reardon, M., Rodriguez, R., Rogers, Y. H., Romblad, D., Ruhfel, B., Scott, R., Sitter, C., Smallwood, M., Stewart, E., Strong, R., Suh, E., Thomas, R., Tint, N. N., Tse, S., Vech, C., Wang, G., Wetter, J., Williams, S., Williams, M., Windsor, S., Winn-Deen, E., Wolfe, K., Zaveri, J., Zaveri, K., Abril, J. F., Guigo, R., Campbell, M. J., Sjolander, K. V., Karlak, B., Kejariwal, A., Mi, H., Lazareva, B., Hatton, T., Narechania, A., Diemer, K., Muruganujan, A., Guo, N., Sato, S., Bafna, V., Istrail, S., Lippert, R., Schwartz, R., Walenz, B., Yooseph, S., Allen, D., Basu, A., Baxendale, J., Blick, L., Caminha, M., Carnes-Stine, J., Caulk, P., Chiang, Y. H., Coyne, M., Dahlke, C., Mays, A., Dombroski, M., Donnelly, M., Ely, D., Esparham, S., Fosler, C., Gire, H., Glanowski, S., Glasser, K., Glodek, A., Gorokhov, M., Graham, K., Gropman, B., Harris, M., Heil, J., Henderson, S., Hoover, J., Jennings, D., Jordan, C., Jordan, J., Kasha, J., Kagan, L., Kraft, C., Levitsky, A., Lewis, M., Liu, X., Lopez, J., Ma, D., Majoros, W., McDaniel, J., Murphy, S., Newman, M., Nguyen, T., Nguyen, N., Nodell, M., Pan, S., Peck, J., Peterson, M., Rowe, W., Sanders, R., Scott, J., Simpson, M., Smith, T., Sprague, A., Stockwell, T., Turner, R., Venter, E., Wang, M., Wen, M., Wu, D., Wu, M., Xia, A., Zandieh, A. & Zhu, X. 2001. The sequence of the human genome. *Science* 291: 1304-1351.
- Vernooij, M. W., Ikram, M. A., Tanghe, H. L., Vincent, A. J., Hofman, A., Krestin, G. P., Niessen, W. J., Breteler, M. M. & van der Lugt, A. 2007. Incidental findings on brain MRI in the general population. *The New England journal of medicine* 357: 1821-1828.
- Vieira, J., Huang, M. L., Koelle, D. M. & Corey, L. 1997. Transmissible Kaposi's sarcoma-associated herpesvirus (human herpesvirus 8) in saliva of men with a history of Kaposi's sarcoma. *Journal of virology* 71: 7083-7087.
- Vitale, F., Viviano, E., Perna, A. M., Bonura, F., Mazzola, G., Ajello, F. & Romano, N. 2000. Serological and virological evidence of non-sexual transmission of human herpesvirus type 8 (HHV8). *Epidemiology and infection* 125: 671-675.
- Vogelstein, B., Fearon, E. R., Hamilton, S. R., Kern, S. E., Preisinger, A. C., Leppert, M., Nakamura, Y., White, R., Smits, A. M. & Bos, J. L. 1988. Genetic alterations during colorectal-tumor development. *The New England journal of medicine* 319: 525-532.
- Vogelstein, B., Papadopoulos, N., Velculescu, V. E., Zhou, S., Diaz, L. A., Jr & Kinzler, K. W. 2013. Cancer genome landscapes. *Science* 339: 1546-1558.
- Warthin, A. S. 1929. Papillary cystadenoma lymphomatosum: a rare teratoid of the parotid region. 13: 116-125.
- Watford, W. T., Moriguchi, M., Morinobu, A. & O'Shea, J. J. 2003. The biology of IL-12: coordinating innate and adaptive immune responses. *Cytokine & growth factor reviews* 14: 361-368.
- Watson, J. D. & Crick, F. H. 1953a. Genetical implications of the structure of deoxyribonucleic acid. *Nature* 171: 964-967.
- Watson, J. D. & Crick, F. H. 1953b. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature* 171: 737-738.
- Weinberg, R. A. 2013. The biology of cancer, second edition. Garland Science, Taylor & Francis Group LLC, and informa business, 711 Third Avenue, NY 10017, USA and 3 Park Square, Milton Park, Abingdon, OX14 4RN, UK.
- Welch, J. S., Ley, T. J., Link, D. C., Miller, C. A., Larson, D. E., Koboldt, D. C., Wartman, L. D., Lamprecht, T. L., Liu, F., Xia, J., Kandoth, C., Fulton, R. S., McLellan, M. D., Dooling, D. J., Wallis, J. W., Chen, K., Harris, C. C., Schmidt, H. K., Kalicki-Veizer, J. M., Lu, C., Zhang, Q., Lin, L., O'Laughlin, M. D., McMichael, J. F., Delehaunty, K. D., Fulton, L. A., Magrini, V. J., McGrath, S. D., Demeter, R. T., Vickery, T. L., Hundal, J., Cook, L. L., Swift, G. W., Reed, J. P., Alldredge, P. A., Wylie, T. N., Walker, J. R., Watson, M. A., Heath, S. E., Shannon, W. D., Varghese, N., Nagarajan, R., Payton, J. E., Baty, J. D., Kulkarni, S., Klot, J. M., Tomasson, M. H., Westervelt, P., Walter, M. J., Graubert, T. A., DiPersio, J. F., Ding, L., Mardis, E. R. & Wilson, R. K. 2012. The origin and evolution of mutations in acute myeloid leukemia. *Cell* 150: 264-278.
- Wiemels, J. L., Cazzaniga, G., Daniotti, M., Eden, O. B., Addison, G. M., Masera, G., Saha, V., Biondi, A. & Greaves, M. F. 1999. Prenatal origin of acute lymphoblastic leukaemia in children. *The Lancet* 354: 1499-1503.

- Wiemels, J., Wrensch, M. & Claus, E. B. 2010. Epidemiology and etiology of meningioma. *Journal of neuro-oncology* 99: 307-314.
- World Health Organization (Editor). 2000. International classification of diseases for oncology, 3rd edition (ICD-O-3). WHO, Geneva, Switzerland.
- Xu, J., Lange, E. M., Lu, L., Zheng, S. L., Wang, Z., Thibodeau, S. N., Cannon-Albright, L. A., Teerlink, C. C., Camp, N. J., Johnson, A. M., Zuhlke, K. A., Stanford, J. L., Ostrander, E. A., Wiley, K. E., Isaacs, S. D., Walsh, P. C., Maier, C., Luedeke, M., Vogel, W., Schleutker, J., Wahlfors, T., Tammela, T., Schaid, D., McDonnell, S. K., DeRycke, M. S., Cancel-Tassin, G., Cussenot, O., Wiklund, F., Gronberg, H., Eeles, R., Easton, D., Kote-Jarai, Z., Whittemore, A. S., Hsieh, C. L., Giles, G. G., Hopper, J. L., Severi, G., Catalona, W. J., Mandal, D., Ledet, E., Foulkes, W. D., Hamel, N., Mahle, L., Moller, P., Powell, I., Bailey-Wilson, J. E., Carpten, J. D., Seminara, D., Cooney, K. A., Isaacs, W. B. & International Consortium for Prostate Cancer Genetics. 2013. HOXB13 is a susceptibility gene for prostate cancer: results from the International Consortium for Prostate Cancer Genetics (ICPCG). *Human genetics* 132: 5-14.
- Xue, W., Kitzing, T., Roessler, S., Zuber, J., Krasnitz, A., Schultz, N., Revill, K., Weissmueller, S., Rappaport, A. R., Simon, J., Zhang, J., Luo, W., Hicks, J., Zender, L., Wang, X. W., Powers, S., Wigler, M. & Lowe, S. W. 2012. A cluster of cooperating tumor-suppressor gene candidates in chromosomal deletions. *Proceedings of the National Academy of Sciences of the United States of America* 109: 8212-8217.
- Zhang, F. & Boothby, M. 2006. T helper type 1-specific Brg1 recruitment and remodeling of nucleosomes positioned at the IFN-gamma promoter are Stat4 dependent. *The Journal of experimental medicine* 203: 1493-1505.
- Zhao, J., Dynlacht, B., Imai, T., Hori, T. & Harlow, E. 1998. Expression of NPAT, a novel substrate of cyclin E-CDK2, promotes S-phase entry. *Genes & development* 12: 456-461.
- Zhao, J., Kennedy, B. K., Lawrence, B. D., Barbie, D. A., Matera, A. G., Fletcher, J. A. & Harlow, E. 2000. NPAT links cyclin E-Cdk2 to the regulation of replication-dependent histone gene transcription. *Genes & development* 14: 2283-2297.
- Zheng, S. L., Sun, J., Wiklund, F., Smith, S., Stattin, P., Li, G., Adami, H. O., Hsu, F. C., Zhu, Y., Balter, K., Kader, A. K., Turner, A. R., Liu, W., Bleecker, E. R., Meyers, D. A., Duggan, D., Carpten, J. D., Chang, B. L., Isaacs, W. B., Xu, J. & Gronberg, H. 2008. Cumulative association of five genetic variants with prostate cancer. *The New England journal of medicine* 358: 910-919.